TWIN CITIES SURFACE WATER SIMULATION MODELING DEMONSTRATION

Special Report 12

**Kevin L. Anderson, John D. Corbett, Nicholas P. Dunning,
Carol A. Gersmehl, Richard P. Greene,
Philip J. Gersmehl, and Dwight A. Brown**

GEOGRAPHY DEPARTMENT

UNIVERSITY OF MINNESOTA

**WATER RESOURCES RESEARCH CENTER
UNIVERSITY OF MINNESOTA
ST. PAUL, MN 55108
JUNE 1987**

TWIN CITIES SURFACE WATER SIMULATION MODELING DEMONSTRATION

Kevin L. Anderson, John D. Corbett, Nicholas P. Dunning, Carol A. Gersmehl, Richard P. Greene, Philip J. Gersmehl, and Dwight A. Brown

ABSTRACT

This report looks at how a water-resources geographic information system (GIS) and computer program can simulate runoff from storm events. We introduce a land-cover classification designed to improved on the way current land-use classifications deal with impervious (paved) surfaces. Our proposed classification also accounts for seasonal changes in vegetative growth and evapotranspiration. The file structure of our GIS uses point observations tabulated within square-kilometer areas. This inventory-based file structure is specifically designed to relate soils and land cover measurements taken from identical places, as well as to identify the range of phenomena within larger areas.

Our study areas consist of two medium-sized watersheds located within the seven-county Twin Cities Metropolitan Area. The Elm Creek basin in northern Hennepin County drains 220 km^2 of urban and suburban developments mixed with agricultural land uses and hobby farms. The Vermillion River basin of central Dakota and eastern Scott Counties (285 km^2) consists primarily of agricultural land, with a number of free-standing communities around which new urban growth is taking place. Both watersheds include a number of land cover types on a diverse mixture of landforms; both also provide an adequate record of stream discharge out of the basin to test simulation results.

The GIS used in this study is the cell-based EPPL7 system that runs on 16-bit personal computers (such as IBM PCs and their compatibles). EPPL7 is a product of the Planning Information Center (PIC), State Planning Agency. The point-counting inventory consists of a series of files, one file for each category in the GIS (e.g., "medium building density" in the land-cover data), with a frequency of occurrence recorded for each cell. We derive land-cover information from Landsat MSS imagery rectified to 100 meter areas on a UTM coordinate base. Soils data are point samples from published county soil surveys. The Appendix shows a method of implementing a point-counting inventories in EPPL7.

Runoff is calculated using the well-known Soil Conservation Service "Curve Number" approach. We produced a series of maps of runoff estimates based on current land-cover from two separate Landsat images for each basin. Estimates are also calculated for an "urbanized" Elm Creek basin to test the applicability of the simulation approach to future growth scenarios.

Three major conclusions can be drawn from this study: (1) useful hydrologic analysis can be performed with a simple GIS and simulation models; (2) the point-counting method is the best approach for coding environmental data; and (3) point-relational data sets are a must for accurate simulation results. Square-kilometer cells are more than adequate for analysis of surface runoff in the Elm Creek and Vermillion River basins. If necessary, results can actually be improved more by better measurements at specific points in an area than by using data at a finer spatial resolution in the GIS.

TWIN CITIES SURFACE WATER SIMULATION MODELING DEMONSTRATION

Kevin L. Anderson, John D. Corbett, Nicholas P. Dunning, Carol A. Gersmehl, Richard P. Greene, Philip J. Gersmehl, and Dwight A. Brown

INTRODUCTION

This report looks at how a water-resources GIS and computer program can be used to simulate runoff from storm events. This report introduces a point-relational strategy for collecting soils and land cover observations from identical places. We discuss how changes to land cover over time affect runoff. These land-cover changes can be of two types: permanent, as in the case of residential construction on previously non-urban land, and temporary, a category that includes seasonal cycles of vegetation growth and crop rotation. We use two medium-sized watersheds within the seven-county Twin Cities Metropolitan Area (TCMA) for our demonstration. Even though our focus is on these fringe areas of the TCMA, our recommendations are general and should have much wider application.

This report is a companion to an earlier study on the application of simulation models to agricultural areas (Anderson et al. 1986; Anderson 1987). That particular study focused on the eighty square-mile Bear Creek watershed outside Rochester in Olmsted County, Minnesota. In Bear Creek, we found that considerable care must be exercised in coding soils data for a GIS to insure identification of widely-scattered soils with high runoff potential. That study emphasized the importance of the timing of storm events and how differences between spring and summer rainfall affect the amount of runoff.

Background

Runoff can vary considerably as a result of changes in land cover that accompany the expansion of cities into the countryside. The general effect of these urban developments is to increase surface runoff for a given rainfall over what it used to be. Newly constructed structures replace formerly vegetated areas, effectively decreasing infiltration of water into the soil. During a storm, water that cannot soak into the soil quickly runs off the surface. Between the storm events, our beautiful bluegrass lawns can be a virtual siphon, transpiring large amounts of water from the soil and tempting us to apply considerable amounts of sprinkler water.

Downtowns and the older residential areas of our cities have been subject to the greatest amount of change in land cover. Almost all of a typical central business district or high density residential area is concrete, asphalt, rooftops, or glass. Future changes in land cover, however, will be minimal since the transition to an urban environment is complete. The area of new development is in the fringe areas that surround the built-up zone. Here, the landscape is undergoing rapid change, and the water resources are susceptible to the effects of policy decisions regarding urban expansion. This zone of landscape change is the setting in which simulation is most valuable to the policy maker.

Water-resource simulation models are abstractions of reality. They can help us estimate the effect of land-cover changes before the development takes place. These estimates can then be used to shape policy decisions that effectively minimize, or at least direct, change to the hydrologic system. Guidelines developed with the assistance of simulation models can help us deal with such timely issues as the preservation of wetlands and water bodies for wildlife and recreation, the measurement and control of changing lake levels, and the development and purification of domestic water sources.

The Twin Cities Metropolitan Area

The seven-county Twin Cities Metropolitan Area represents the largest, most populated area in Minnesota. It provides the biggest laboratory available to assess the use of GIS with simulation models in an urbanizing environment (Figure 1). Beyond the urban core (Minneapolis, St. Paul, and their neighboring suburbs) is a fringe zone that is developing at a steady pace. Apple Valley, Brooklyn Park, Burnsville, Champlin, Eagan, Eden Prairie, Mendota Heights, Shoreview, Vadnais Heights, and Woodbury are just a few of the communities involved. By the year 2000, the TCMA will have an estimated population of 2.3 to 2.9 million people, an increase of 15 to 25 percent over the nearly 2.0 million people already living in the region in 1980 (U.S. Bureau of the Census 1983; Metropolitan Council of the Twin Cities 1977; Upper Mississippi River Basin Commission 1977). Almost all of this new growth will be within the developing fringe.

The Twin Cities Metropolitan Area is physically large. Data collection for the area is not a trivial task. If you draw a box around the TCMA on a map, the region is 64 miles on a side -- roughly 4,000 square miles (10,000 square kilometers). A good water-resources GIS will have data from every one of those miles in order to provide complete coverage. The number of sources one must consult for the most common data elements is staggering (Table 1).

TABLE 1: SIZE OF THE TWIN CITIES METROPOLITAN AREA AND COVERAGE BY MOST COMMON DATA SOURCES AND SCALES

10,000	km^2 (1 million hectares)
4,000	mi^2 (64,000 40-acre parcels)
435	Detailed Soil Survey Sheets (at 1:20,000 or 1:15,840 scale)
221	Large-Scale Aerial Photographs (at 1:9,600 scale)
77	Topographic Maps (at 1:24,000 scale)

The TCMA is physically complex. The area contains fourteen different geomorphic regions, contributes to three of the state's principal rivers and many more minor streams , is served by at least three major aquifers, and contains well over 100 lakes and water bodies. The region has a full range of land uses, from productive farmland to the most densely built urban land. A GIS developed for this region must be able to inventory all possible combinations of this far-from-simplistic landscape at a scale detailed enough for simulation.

Such a water-resources GIS currently does not exist for the TCMA. Pre-collected data appropriate for hydrologic analysis are not yet available for the

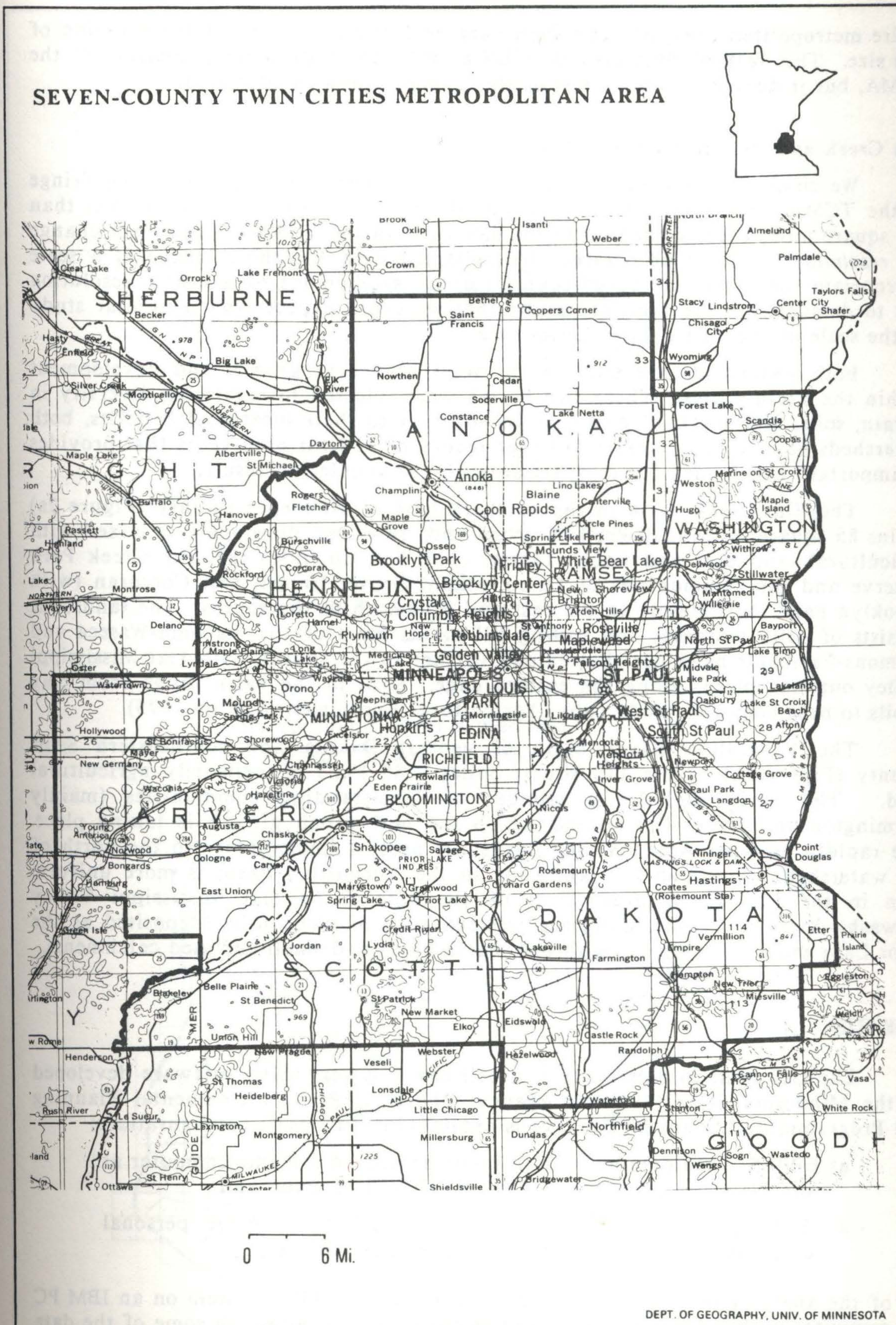

SEVEN-COUNTY TWIN CITIES METROPOLITAN AREA
SHERBURNE
ANOKA
WASHINGTON
RAMSEY
HENNEPIN
CARVER
SCOTT
DAKOTA
MINNEAPOLIS
ST PAUL
Anoka
Coon Rapids
Brooklyn Park
Brooklyn Center
Fridley
White Bear Lake
Roseville
Maplewood
Golden Valley
Robbinsdale
ST LOUIS PARK
EDINA
RICHFIELD
BLOOMINGTON
MINNETONKA
Hopkins
West St Paul
South St Paul
Shakopee
Stillwater
Hastings
Forest Lake
Elk River
Farmington
Lakeville
Chaska
0 6 Mi.
DEPT. OF GEOGRAPHY, UNIV. OF MINNESOTA

FIGURE 1

entire metropolitan area, nor can such data be collected overnight for a region of this size. This lack of data may prohibit a full-scale hydrologic simulation of the TCMA, but it does still allow a test of the "concept" in a smaller area.

Elm Creek and Vermillion River Basins

We chose to focus on two small watersheds found in the developing fringe of the TCMA. Even though we changed the scale of analysis to areas less than 150 square miles each (compared to the 4,000 mi^2 in the TCMA), we did not change the *resolution* (or level of detail). We collected data for those basins at a scale appropriate for the entire seven-counties area. Our choice of a coarse resolution was to demonstrate the kind of results that we could expect from a similar study on the scale of the entire Twin Cities area.

Both watersheds are adjacent to areas of recent growth and development within the Twin Cities. These two basins were selected to maximize diversity in terrain, soils, and surficial geology. In addition to their locational benefits, both watersheds have a daily stream disharge record of at least nine years that provides an important benchmark for comparison with our simulation results.

The Elm Creek basin in the northern half of Hennepin County (Figure 2a) drains 85 square miles (220 km^2) of urban and suburban developments mixed with agricultural land uses and hobby farms. The area includes the Elm Creek Park Reserve and the communities of Osseo, Maple Grove, Hamel, and Corcoran (with Brooklyn Park immediately to the east and Plymouth to the south). The landscape consists of hilly moraines and rolling till plains (part of the Waconia-Waseca and Emmons-Faribault moraines that extend farther to the south), with flat Mississippi Valley outwash plains along the eastern edge. Soils in the basin range from clay knolls to peat bogs and sandy fans (Agricultural Experiment Station 1975).

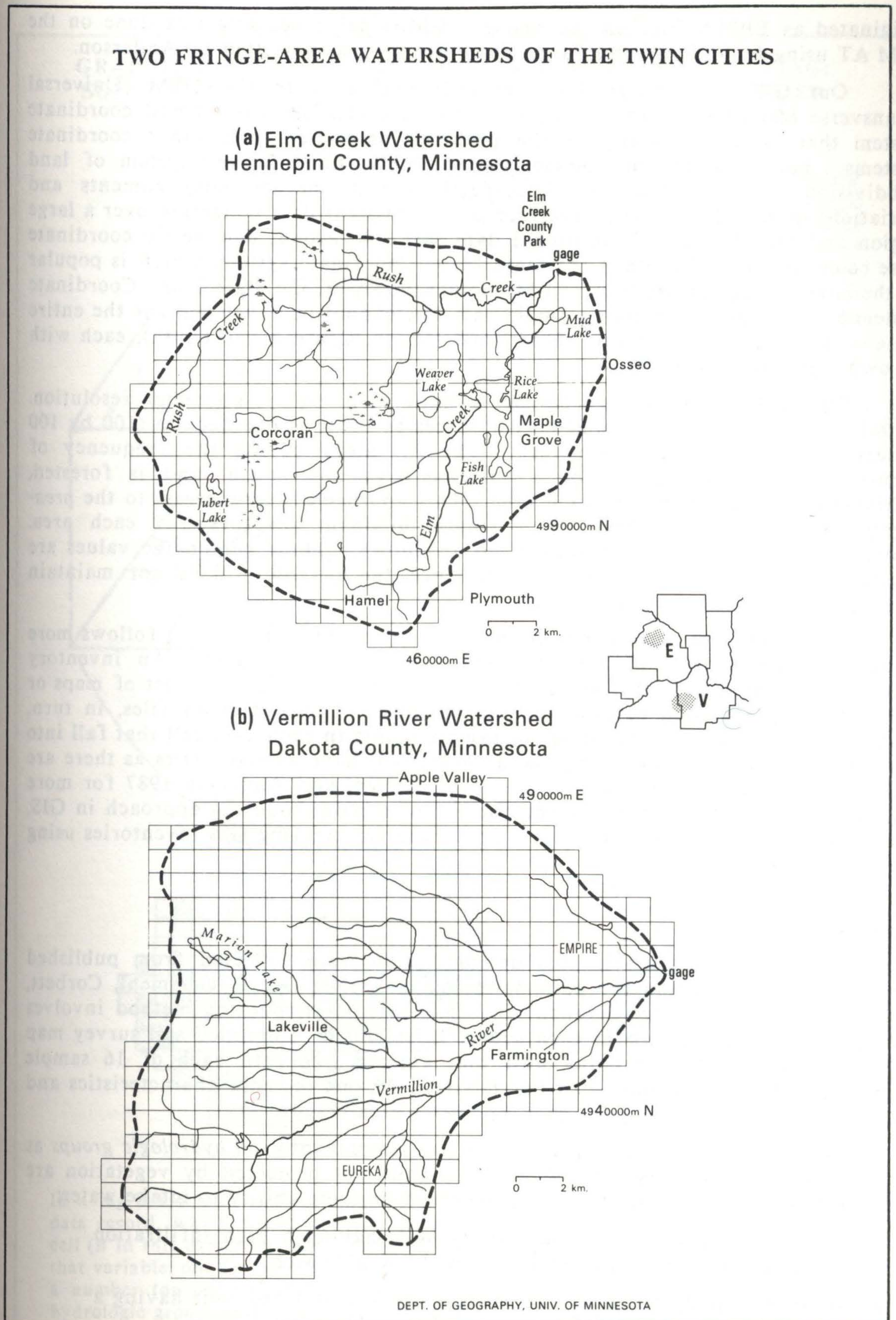
The Vermillion River basin of central Dakota and southeastern Scott County (Figure 2b) covers 110 square miles (285 km^2) of primarily agricultural land. The basin includes a number of free-standing communities (mainly Farmington and Lakeville) around which the new urban growth is taking place. The rapidly growing Apple Valley-Burnsville area is immediately to the north of the watershed. Although the landscape in the Vermillion Basin is more uniform than in the Elm Creek area (most of the area is on flat Mississippi Valley outwash), there is a zone along the western edge of steep hills (deposited as the Prior Lake moraine complex) with low marsh areas and small enclosed catchments.

THE GIS

The GIS used in this study is a cell-based system using software developed by the Minnesota State Planning Agency. Their EPPL (Environmental Planning and Programming Language) package is available in two basic configurations:

- as EPPL6, a mainframe version implemented on a Prime computer at the Planning Information Center (PIC) in St. Paul, and
- as EPPL7, a newer microcomputer version for 16-bit personal computers (such as IBM PCs and their compatibles).

All of the analyses in this report were done using the EPPL7 system on an IBM PC AT computer with a Professional Graphics color display, although some of the data



TWO FRINGE-AREA WATERSHEDS OF THE TWIN CITIES
(a) Elm Creek Watershed
Hennepin County, Minnesota
Elm Creek County Park
gage
Rush
Creek
Mud Lake
Osseo
Weaver Lake
Rice Lake
Maple Grove
Corcoran
Fish Lake
Jubert Lake
Elm
490000m N
Hamel
Plymouth
0
2 km.
460000m E
E
V
(b) Vermillion River Watershed
Dakota County, Minnesota
Apple Valley
490000m E
Marion Lake
EMPIRE
gage
Lakeville
River
Farmington
Vermillion
494000m N
EUREKA
0
2 km.
DEPT. OF GEOGRAPHY, UNIV. OF MINNESOTA

FIGURE 2

originated as EPPL6 files on the Prime. Additional processing was done on the IBM AT using FORTRAN and Pascal computer programs written by Anderson.

Our GIS uses square-kilometer cells registered to the UTM (Universal Transverse Mercator) coordinate base. UTM is a flexible, earth-based coordinate system that minimizes many of the problems associated with other coordinate systems. For instance, the commonly-used Public Land Survey system of land subdivision has a number of imperfections (e.g., border misalignments and variations in parcel sizes) that interfere with consistent data collection over a large region and the merging of additional data sets. A second, non-metric coordinate base commonly used in GIS is the State Plane Coordinate system, which is popular in the engineering community. In Minnesota, however, the State Plane Coordinate system has the basic flaw that a single base file cannot be developed for the entire state -- the area is divided into three zones (North, Central, and South), each with its own origin of reference.

Our GIS consists of *point-count* inventories at square-kilometer resolution. Detailed observations were collected for places as small as one hectare (100 by 100 meters or approximately 2.5 acres). The inventory counts the frequency of occurrence of each category of a variable (e.g., how much land is forested, cultivated, and urban) within each area. This approach is in contrast to the *area-tagging method*, which identifies a single dominant category for each area. Percentage values are used to record our inventory. These percentage values are similar to a *probability of occurrence*, since the inventory does not maintain locational information within the cell.

The EPPL7 software is not designed as an inventory system (it follows more traditional GIS method of area-tagging by dominant features). An inventory system can be approximated, however, by coding each variable as a set of maps or files, one map for each category of the variable. These category files, in turn, consist of records of the percentage of sample points in each data cell that fall into that category. Each variable (e.g., land cover) will have as many files as there are categories for the variable (See P. Gersmehl, Brown, and Anderson 1987 for more discussion on coordinate systems and the use of a point-counting approach in GIS; the Appendix in this report outlines a procedure for creating GIS inventories using the EPPL7 system).

Soils Data

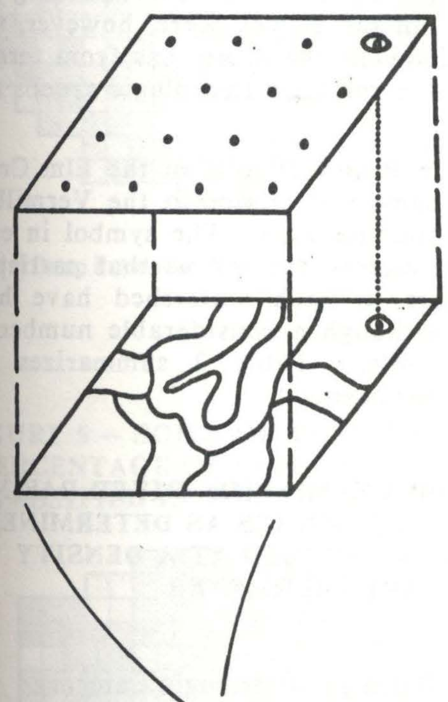
The soils data used in this demonstration were collected from published county soil surveys using the point-sampling technique (see P. Gersmehl, Corbett, and Greene 1987 for more details). In brief, this point-counting method involves laying a transparent grid over each square kilometer of a detailed soil survey map and recording the soil mapping unit that appears beneath each of 16 sample points. These mapping units can be classified by one or more characteristics and summarized as an inventory for each kilometer area (Figure 3).

For this study, we "interpreted" the soil samples into *soil hydrologic groups* as defined by the Soil Conservation Service. Soils not protected by vegetation are assigned to one of four hydrologic groups based on their ability to intake water:

Group A -- Deep, well drained sandy soils having a high infiltration rate (low runoff potential) when thoroughly wet;

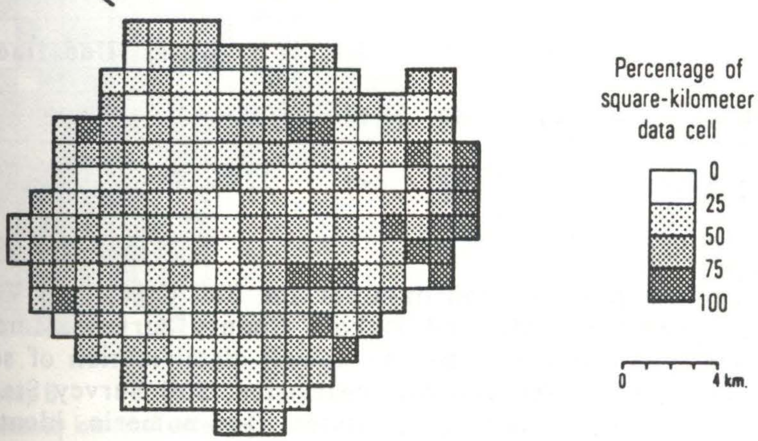
Group B -- Moderately deep or moderately well drained soils having a moderate infiltration rate when thoroughly wet;

GRAPHIC DISPLAY OF DATA OBTAINED BY POINT-COUNTING: SOIL HYDROLOGIC GROUPS IN THE ELM CREEK WATERSHED



This example illustrates the method for obtaining an inventory of soil hydrologic groups within a single square-kilometer data cell. A transparent template is placed on the detailed soil map, and the soil mapping unit beneath each dot on the template is recorded. The next step is to reclassify the taxonomic mapping units into soil hydrologic groups; the table shows the estimated percentage of the soils in the data cell that belong in each hydrologic group.

Hydrologic Soil Group	Number of Samples	Percentage Occurrence
A	1	6
B	9	56
C	2	13
D	4	25
TOTAL	16	100



In a tag-based GIS, the variable "Soil Hydrologic Group" would contain a single data record, which would contain the letter of the dominant group in each data cell (B in this example). A count-based GIS, by contrast, would have four files for that variable, one for each hydrologic group, and each of those records would have a number (or percentage) that reflects the proportion of the data cell in that hydrologic group, as determined by the point-counting inventory in that cell.

FIGURE 3

Group C -- Fine texture soils having a slow infiltration rate when wet and a layer that impedes downward movement of water; or

Group D -- Fine textured soils with a high water table or shallow impervious layer and having a very slow infiltration rate (high runoff potential) when thoroughly wet (Soil Conservation Service 1972, p. 7.2).

Several soil mapping units (urban land, complex soils) do not have an assigned soil hydrologic group. These unassigned soils were ignored in our sample, with the sample size in each kilometer adjusted downward to represent the remaining soil inventory. Point samples designated as "water" in the survey were, however, still included in our inventory. Each individual kilometer, therefore, has from zero to sixteen samples comprising its inventory of soils in the four hydrologic groups and water.

Figure 4 presents the soil data for Group B and D soils in the Elm Creek basin; Figure 5 presents the inventory for the same soil groups in the Vermillion River basin. Each square represents a square-kilomter area. The symbol in each map square indicates the percentage of the kilometer observed as that particular hydrologic group. The majority of kilometers in each watershed have high percentages or high counts for Group B soils, although a considerable number of kilometers contain poorly-drained Group D soils. Table 2 summarizes the differences between the soil inventories of the two basins.

TABLE 2: COMPARISON OF ELM CREEK AND VERMILLION RIVER BASINS: PERCENTAGE OF AREA IN HYDROLOGIC CATEGORIES AS DETERMINED FROM DETAILED COUNTY SOIL SURVEYS SAMPLED AT A DENSITY OF 16 DATA POINTS PER SQUARE KILOMETER

	Percentage of Basin in Hydrologic Category					
	A	**B**	**C**	**D**	**Water**	**Undefined**
Elm Creek	<1	49	13	28	2	8
Vermillion River	4	69	<1	22	3	2

Three county surveys provided soil data for the Elm Creek and Vermillion River basins. The Hennepin County and Dakota County surveys (Lueth 1974; Hundley 1983) are modern surveys that use the newest classification of soil series (known as the Soil Taxonomy or Seventh Approximation; Soil Survey Staff 1975). The Hennepin County survey, however, predates the numeric identification procedure. The Scott County survey (Harms 1959) is an old survey that was completed long before the use of contemporary Soil Taxonomy and should be brought up-to-date. A number of soil characteristics normally described in modern surveys are missing or out-of-date in the Scott County survey. Many of the soil series identified for the county have since been redefined, some into multiple series with narrower ranges of traits than the broad definitions in the Scott County survey. Most of the land was agricultural when the survey was made, which is not the case today. Although we were able to use the Scott County survey for this

FIGURE 4 -- SOILS OF THE ELM CREEK BASIN: PERCENTAGE OF SAMPLE POINTS IN EACH DATA CELL CLASSIFIED IN HYDROLOGIC GROUPS B AND D

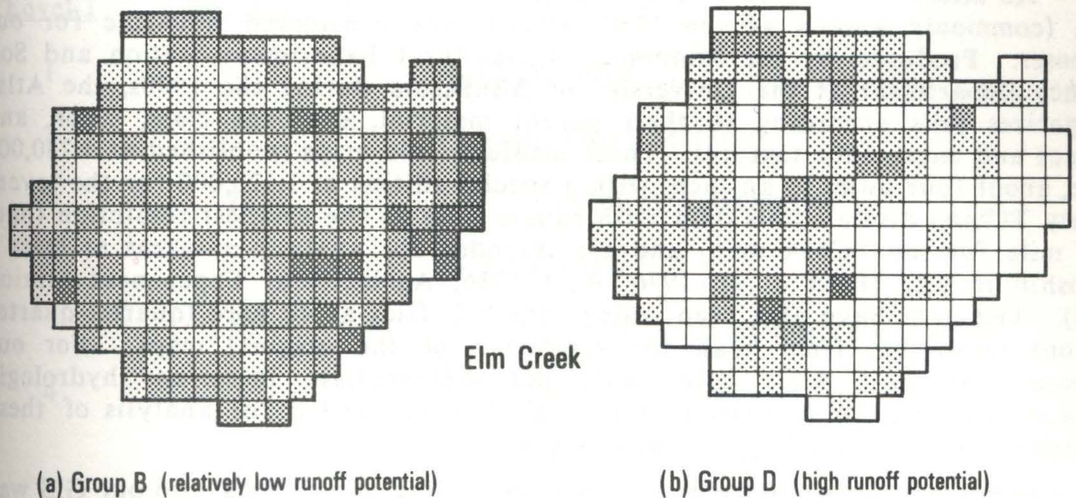
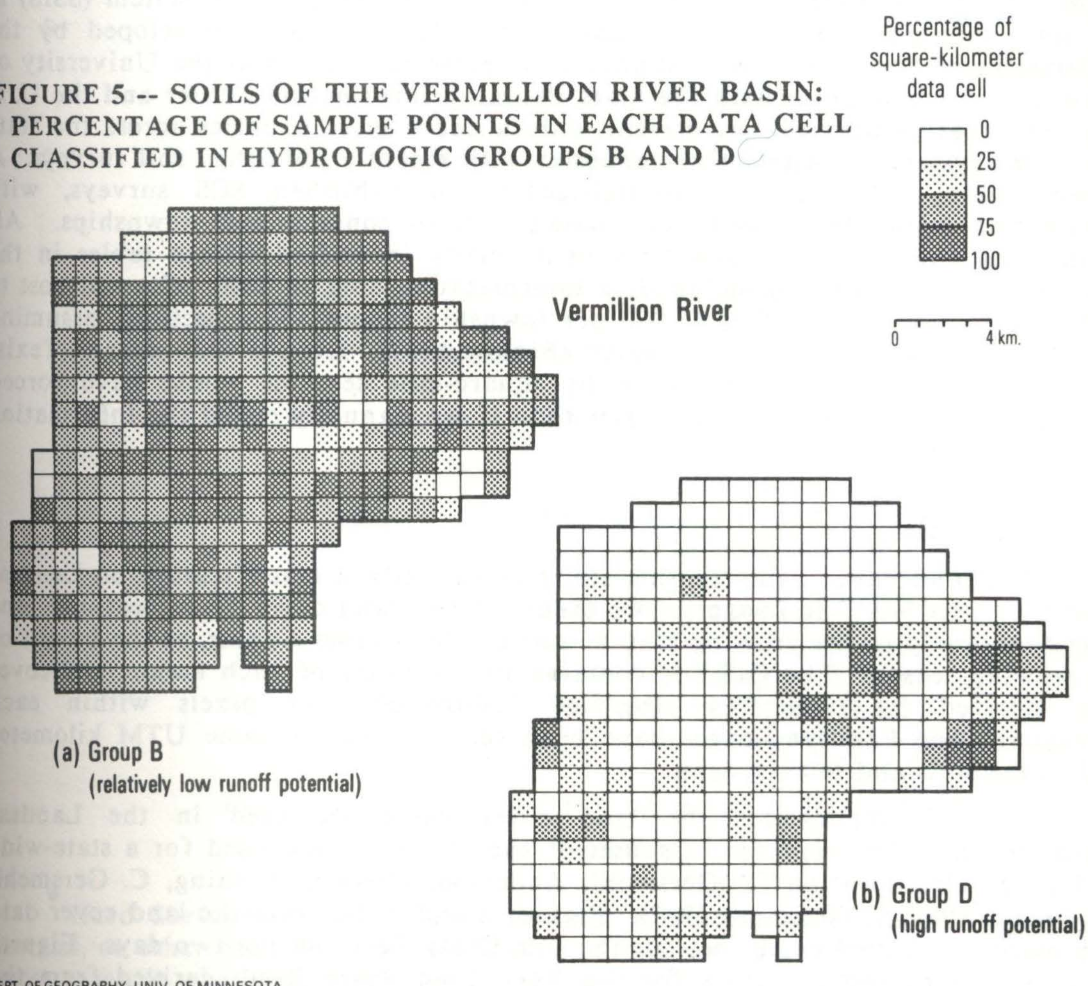


FIGURE 5 -- SOILS OF THE VERMILLION RIVER BASIN: PERCENTAGE OF SAMPLE POINTS IN EACH DATA CELL CLASSIFIED IN HYDROLOGIC GROUPS B AND D



DEPT. OF GEOGRAPHY, UNIV. OF MINNESOTA

project (the amount of the Vermillion River basin in the county is small), the accuracy of information for that county should be viewed as suspect.

An alternative source of soil information, the state-wide *Soil Landscape* map data (commonly known as the "Soil Atlas"), was considered unusable for our purposes. Produced by the Minnesota Agricultural Experiment Station and Soil Science Department at the University of Minnesota during the 1970s, the Atlas categorizes soils according to their parent material, drainage class, color, and surface and subsurface textures. These medium-scale maps (published at 1:250,000 scale, about four miles to an inch, with a special edition at 1:125,000 for the seven-county TCMA) display soils data with square mile as the smallest resolution (one-half mile for the TCMA map) and are intended for analysis of areas at least a township in size (Rust et al. 1976; Rust 1986; Agricultural Experiment Station 1975). The data have also been coded into PIC files at 40 acre (quarter-quarter section) resolution, finer than the resolution of the original maps. For our purposes, this Soil Atlas data could not differentiate important hydrologic characteristics known to exist at the resolution we needed for analysis of these watersheds (P. Gersmehl, Corbett, and Greene 1987).

In passing, we should note that manual coding of soil data into our GIS was necessary, because none of the seven counties within the TCMA had their detailed county surveys digitized into the state's Soil Survey Information System (SSIS) at the time this study was started. SSIS is an electronic atlas developed by the Department of Soil Science and Minnesota Experiment Station at the University of Minnesota in cooperation with the State Planning Information Center and the U. S. Soil Conservation Service. At present, only a dozen or so counties across the state are in this system, primarily those counties with recently published soil surveys. A county survey in SSIS is scan-digitized from published SCS surveys, with information stored by square-mile sections within congressional townships. All engineering and physical characteristics normally identified within tables in the published surveys are also included as interpretative data in SSIS. Current cost to code information into SSIS is $1,000 per township (about $30 a section), assuming that the published survey already exists (Finney 1987). Even if SSIS data did exist for our watersheds, its organization by square-mile sections would have forced resampling of the data at a cost greater than if we manually coded the information from soil surveys.

Land Cover Data

The land cover data for this study is basically a classification of Landsat remotely-sensed satellite imagery. We produced two land cover classifications: one from June, early in the growing season, and a second from August, at the height of the growing season. The GIS file contains an inventory of each major land-cover type, derived by tabulating classified 100-by-100-meter pixels within each kilometer. The Landsat images have been rectified to the same UTM kilometer grid as the soils and terrain data.

Table 3 summarizes the land cover categories used in the Landsat classification. They represent a subset of the categories proposed for a state-wide land cover classification (P. Gersmehl, Anderson, Greene, Dunning, C. Gersmehl, and Brown 1987; C. Gersmehl 1987). Figures 6 and 7 illustrate the land cover data with maps of selected categories for the Elm Creek Basin on the two days. Figures 8 and 9 present the inventory for the Vermillion River Basin derived from the same imagery. Differences in land cover between the two basins and time periods are summarized in Table 4.

TABLE 3: CATEGORIES USED WITH THE LANDSAT CLASSIFICATION OF TWIN CITIES LAND COVER

Level I	Level II	Category
1		Hard Surfaces*
2		Pervious Earth Materials
	21	Gravel pits; extractive
3		Surface Water
	30	Surface Water (Undifferentiated)
4		Persistent Vegetation
	41	Forested; woods; trees
	42	Grassland (Pasture, prairie, open-space recreation)
5		Wetlands
	50	Wetlands (Undifferentiated)
6		Temporarily Vegetated Areas
	61	Cover Crop (Hay, Alfalfa)
	62	Small Grains
	63	Row Crops (Corn, Beans)
7		Developed Areas
	71	Commercial/Industrial Building Density (large structures, +85 % impervious)
	72	High Building Density (HBD) (<1/8 acre, small to medium-sized structures, roughly 65 % imp.)
	73	Medium Building Density (MBD) (1/8 to 1/2 acre, small structures, roughly 35 % imp.)
	74	Low Building Density (LBD) (1/2 to 1 acre, small structures, roughly 25 % imp.)
	75	Very Low Building Residential (VBD) (>1 acre, small structures, roughly 20 % imp.)

*"Hard Surfaces" as a category was not distinquishable from commercial, industrial, or other high density land covers using Landsat at 100-meter resolution; in most cases, it was lumped in with Category 71, "Commercial/Industrial Building Density".

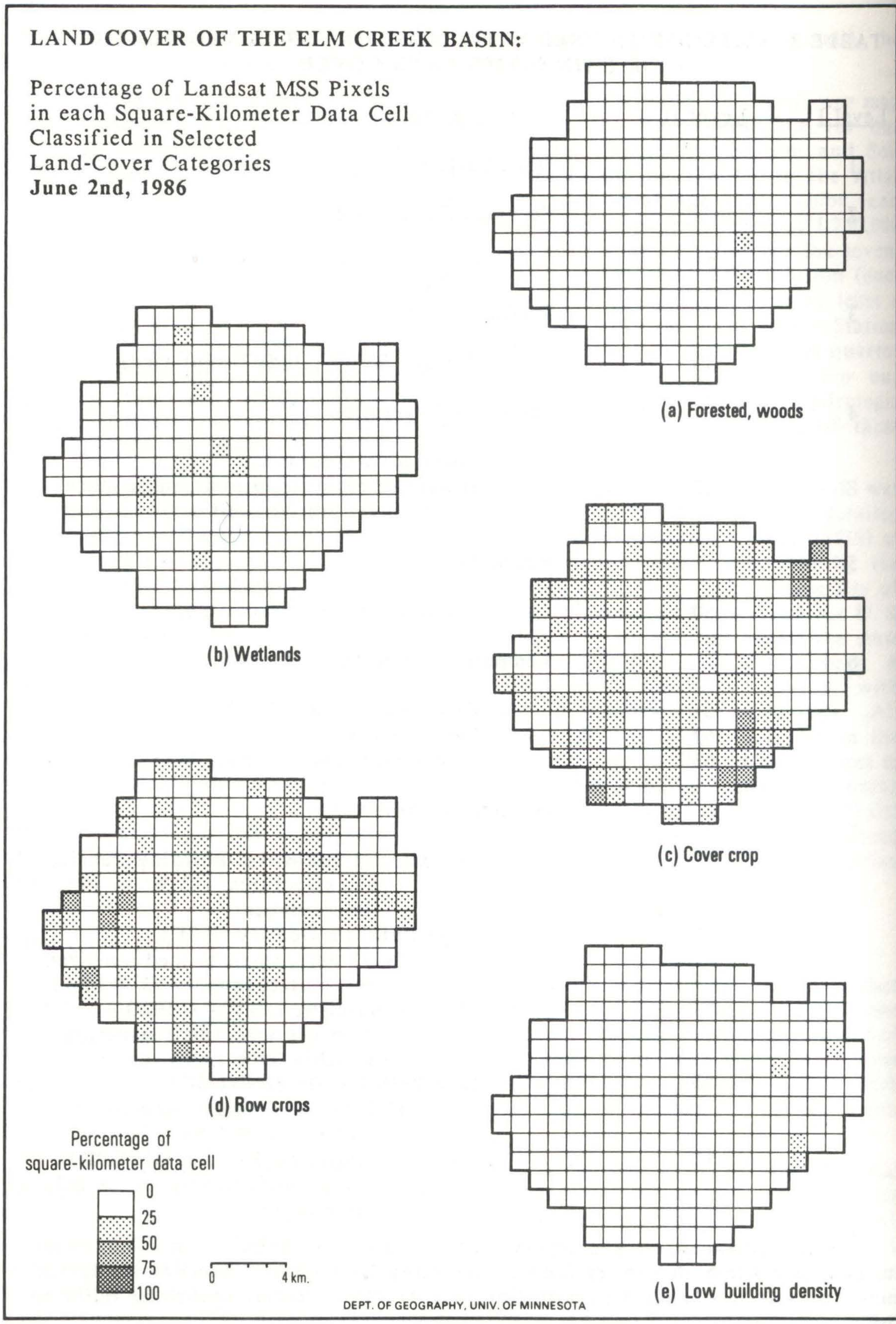

LAND COVER OF THE ELM CREEK BASIN:
Percentage of Landsat MSS Pixels
in each Square-Kilometer Data Cell
Classified in Selected
Land-Cover Categories
June 2nd, 1986
(a) Forested, woods
(b) Wetlands
(c) Cover crop
(d) Row crops
(e) Low building density
Percentage of
square-kilometer data cell
0
25
50
75
100
0
4 km.
DEPT. OF GEOGRAPHY, UNIV. OF MINNESOTA

FIGURE 6

LAND COVER OF THE ELM CREEK BASIN:

Percentage of Landsat MSS Pixels
in each Square-Kilometer Data Cell
Classified in Selected
Land-Cover Categories
August 21st, 1986

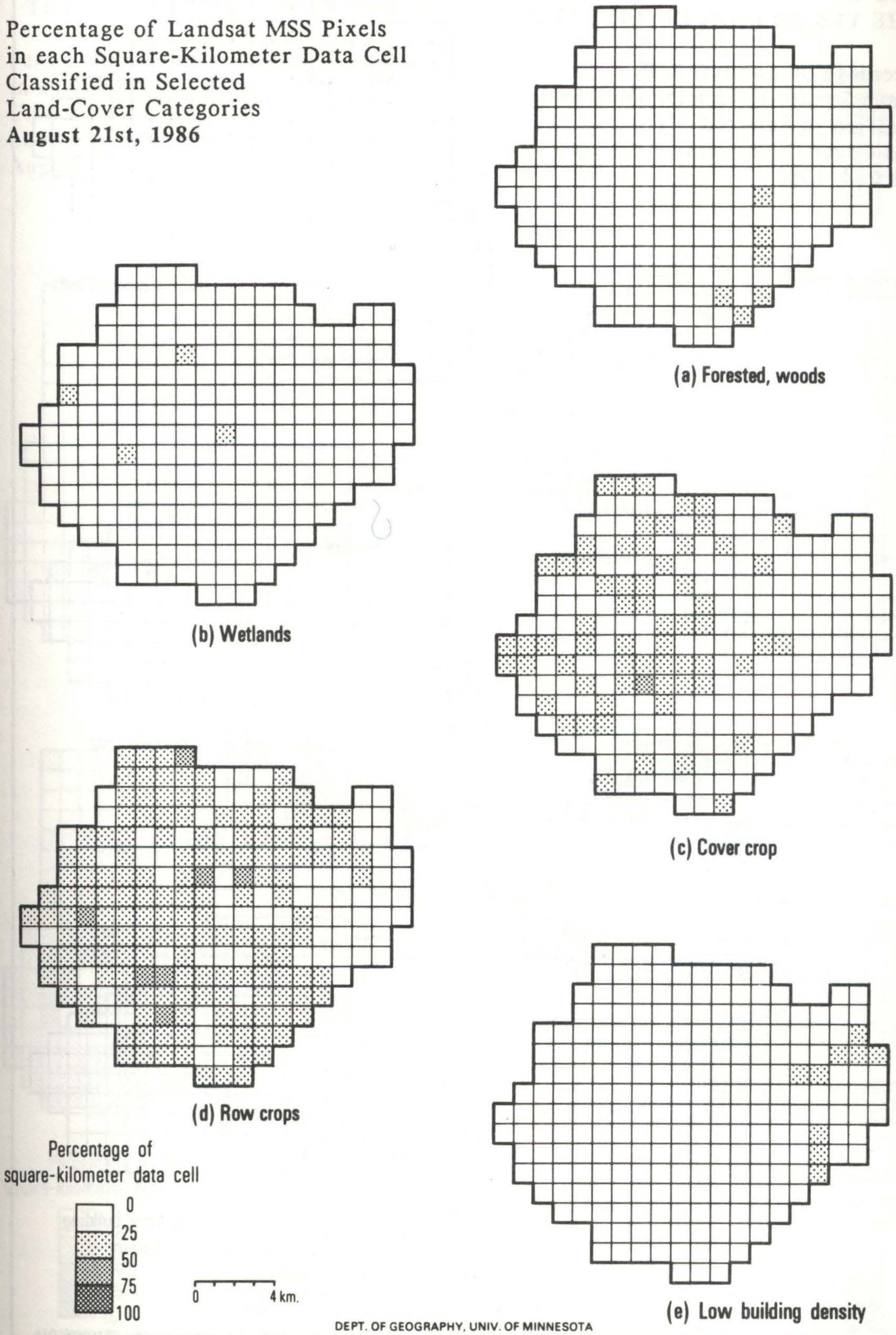


FIGURE 7

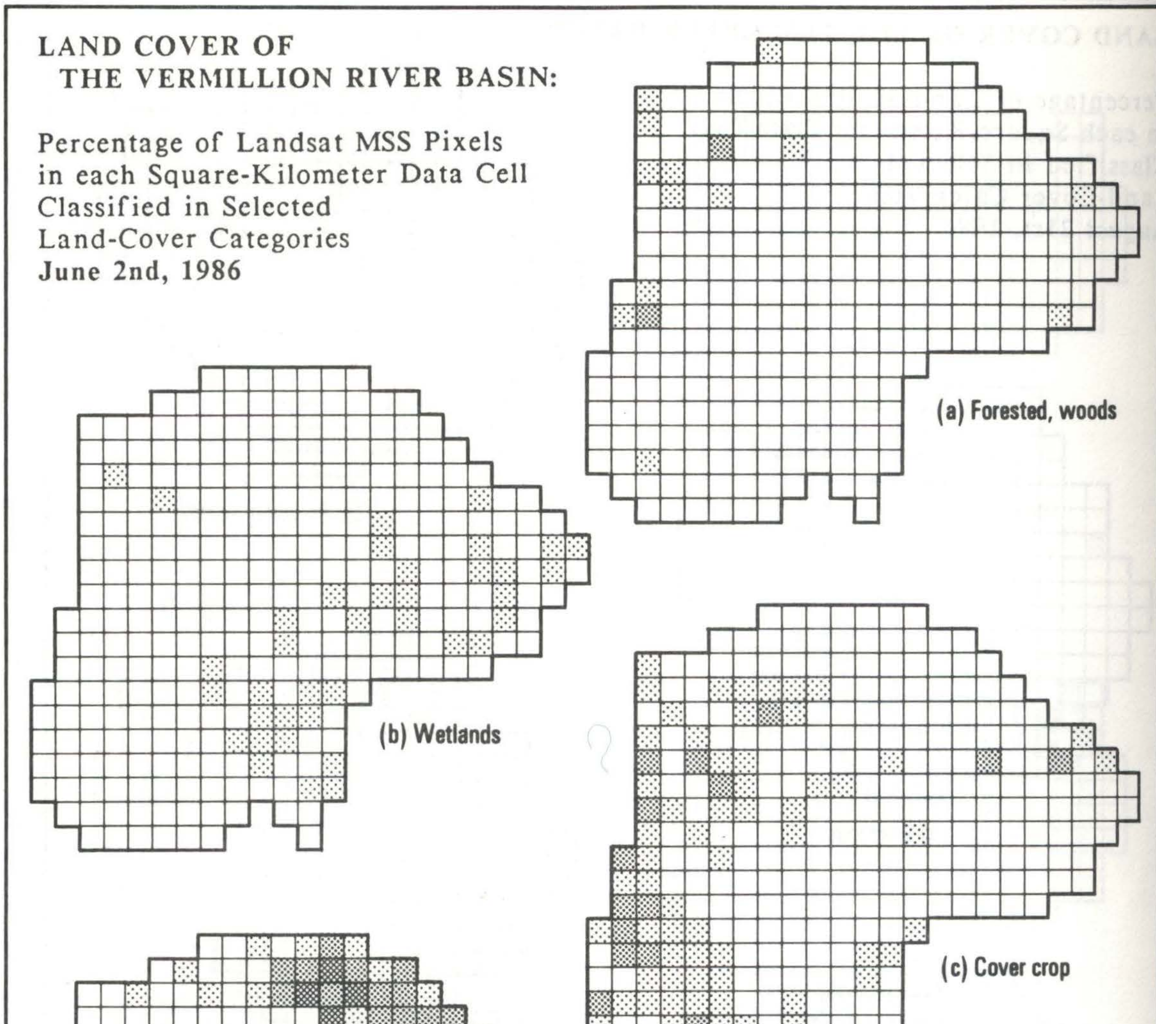

LAND COVER OF
THE VERMILLION RIVER BASIN:
Percentage of Landsat MSS Pixels
in each Square-Kilometer Data Cell
Classified in Selected
Land-Cover Categories
June 2nd, 1986
(a) Forested, woods
(b) Wetlands
(c) Cover crop
(d) Row crops
(e) Low building density
Percentage of
square-kilometer data cell
0
25
50
75
100
0
4 km.
DEPT. OF GEOGRAPHY, UNIV. OF MINNESOTA

FIGURE 8

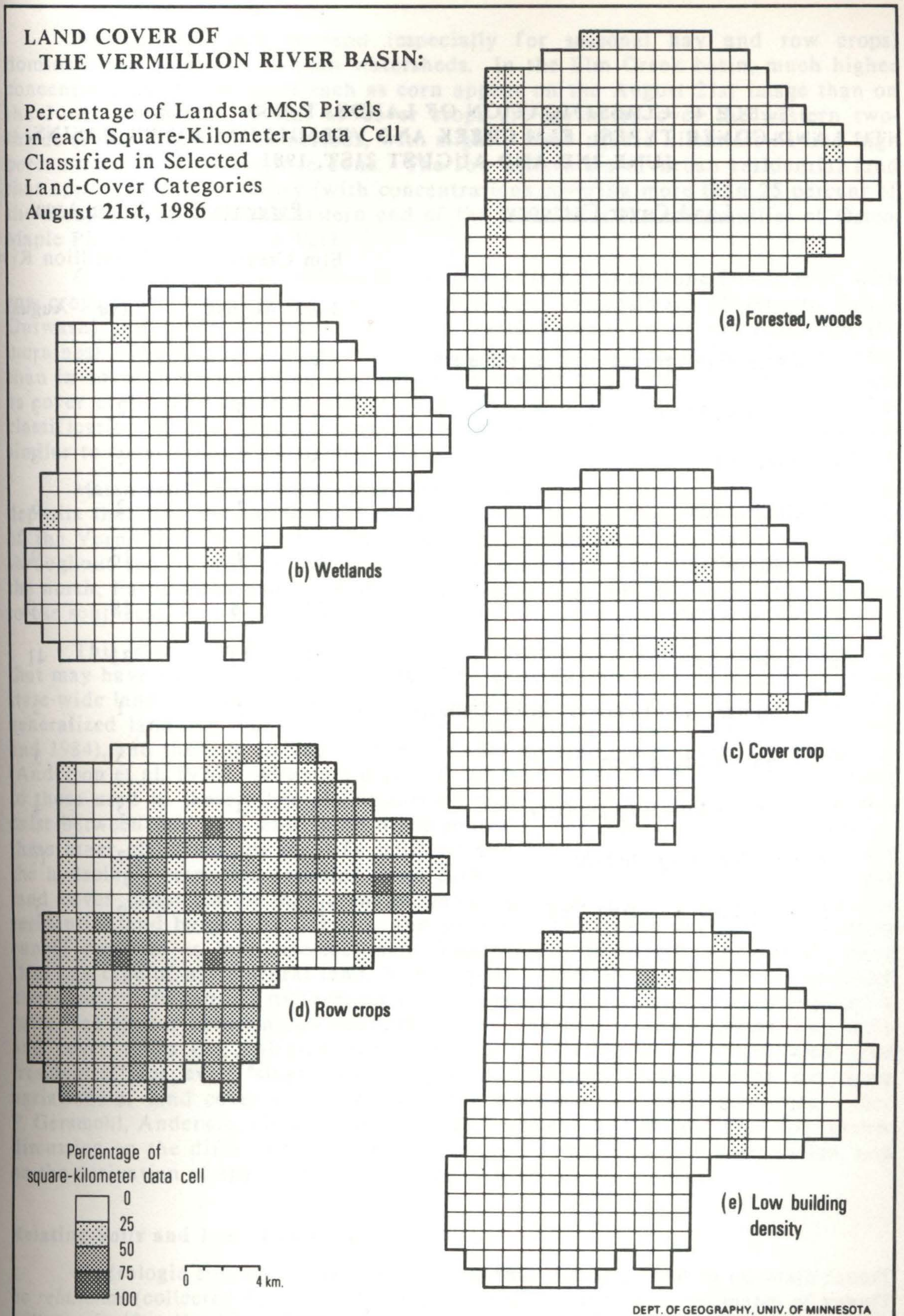

LAND COVER OF
THE VERMILLION RIVER BASIN:
Percentage of Landsat MSS Pixels
in each Square-Kilometer Data Cell
Classified in Selected
Land-Cover Categories
August 21st, 1986
(a) Forested, woods
(b) Wetlands
(c) Cover crop
(d) Row crops
(e) Low building density
Percentage of
square-kilometer data cell
0
25
50
75
100
0
4 km.
DEPT. OF GEOGRAPHY, UNIV. OF MINNESOTA

FIGURE 9

TABLE 4: CLASSIFICATION OF LANDSAT MSS IMAGES INTO LAND-COVER TYPES: ELM CREEK AND VERMILLION RIVER BASINS, JUNE 2ND AND AUGUST 21ST, 1981

Land Cover Category		Percentage of Basin Area			
		Elm Creek		Vermillion River	
		June	**August**	**June**	**August**
21	Extractive (gravel pits, dirt roads)	1	1	1	1
30	Surface Water	2	3	2	3
41	Forested	3	7	6	8
42	Grasslands	3	7	2	6
50	Wetlands	11	9	16	9
61	Cover Crop	27	18	19	10
63	Row Crop	23	30	33	41
71	Commercial/Industrial Bldg. Density	2	2	1	2
72	High Building Density	4	3	3	3
73	Medium Building Density	5	4	5	4
74	Low Building Density	9	11	7	10
75	Very Low Building Density	11	7	6	5

Agricultural uses of land (especially for seasonal hay and row crops) dominate the landscape in both watersheds. In the Elm Creek basin, much higher concentrations of row crops such as corn appear on the August 21st image than on the June 2nd image. Fields of cover crops are sprinkled over the western two-thirds of the area in both periods, with slightly more square kilomters having high percentages of cover crops in June. The few kilometers of urban residential land that appear in the inventory (with concentrations covering more than 25 percent of the area) are found in the eastern end of the basin near the communities of Osseo, Maple Plain, and Brooklyn Park.

Cropland in the Vermillion River basin (Plate 7-4) on June 2nd is split, with row crops concentrated in the eastern two-thirds of the basin (on Mississippi Valley Outwash) and cover crops located in the western one-third of the basin (on the moraines). The August 21st land cover data indicates even more land in row crops than in cover crop, including the southwestern edge of the basin earlier identified as cover crop. The apparent reduction in cover crop may indicate a possible mis-classification of the Landsat image into categories that have spectral signatures similar to cover crops (C. Gersmehl 1987).

Marsh lands have been identified on the June 2nd image near the alluvial deposits that line the shallow valley floor along the southern and eastern reaches of the Vermillion River. As with Elm Creek, urban residential lands are spread throughout the Vermillion basin, with some concentrations near Apple Valley to the north, Farmington and Lakeville in the central part of the basin, and Hampton to the southeast.

This *land cover* data has a number of advantages over the existing data sets that may have been used for this study. The existing data sources include the 1969 state-wide land use map of Minnesota (Minnesota State Planning Agency 1969), the generalized land use maps for the Twin Cities (Metropolitan Council 1978, 1980, and 1984), and the U.S. Geological Survey land use/land cover inventory map series (Anderson et al. 1976). Although many of our land cover categories can be reduced to those used in these other classifications, some important hydrologic differences exist between our data and those other sources. The *land use* categories used in these other classfications identify political and economic differences of land, not the hydrologic responses expected from each surface condition. For instance, our land cover classification has a major category, "developed areas", that separates residential land by density, lot size, and pavement to capture important changes in runoff response as land becomes more impervious. Similarly, we identify more than one type of agricultural land, "temporarily vegetated surface" in our proposed classification, to capture differences in the vegetative cover and evapotranspiration rates that appear between row crops (e.g., corn and beans), small grains (e.g., oats), and cover crops (e.g., alfalfa and other hay). Categories of "cultivated" and "residential" (or even "single-family" and "multi-family") can include too many variations of land cover to be useful in the assessment of water resources. (See P. Gersmehl, Anderson, Greene, Dunning, C. Gersmehl, and Brown 1987 for further discussion on the differences between a land-*cover* and land-*use* classification, and on the derivation of the land cover inventory used in this report.)

Relating Soils and Land Cover data

Hydrologic simulation usually requires that the data used to estimate runoff be *relational* (collected for the same point in space). Incorrect estimates of runoff will result if soils data is collected from one location, land cover data from a second place, and slope measurements from a third.

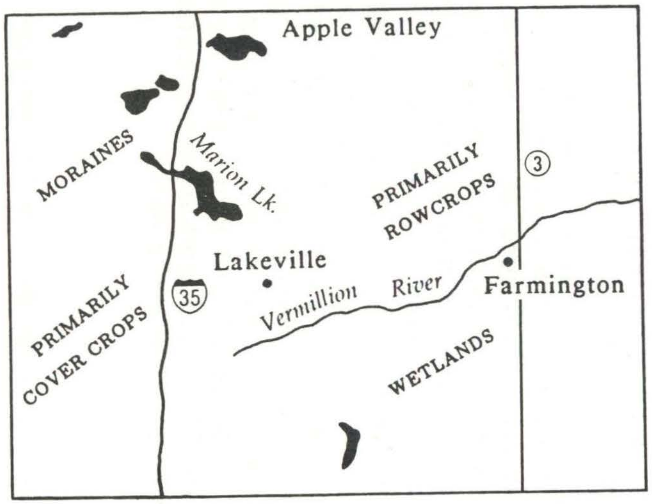


Plate 7-4. Color photographs of June (A) and August (B) rectified Landsat images of the Vermillion River basin of western Dakota and eastern Scott Counties show how the two time periods appear on a computer monitor. On both images, lakes are black because water has low reflectance on all four MSS bands. The moraines along the western edge of the basin appear as bright-reddish hues (indicating healthly grass and trees) in the **June** image and as dark (nearly black) colors on the **August** image. The blues and greens on the June image are bare fields, urban developments, or wetlands; the reds on the August image are primarily cropland. The urban cover extends across the northern edge of the basin in the Apple Valley area, south along the interstate, and clusters around Lakeville and Farmington.

We used nearly 50 **training areas** (i.e., areas whose land cover was known) in order to classify the June and August images into hydrologically significant land-cover categories. The color photograph of the August classification (C) illustrates the general arrangement of land covers. The classification colors have the following meanings:

Blues and aqua - water and wetlands

Pink - high building densities

Red - medium building densities

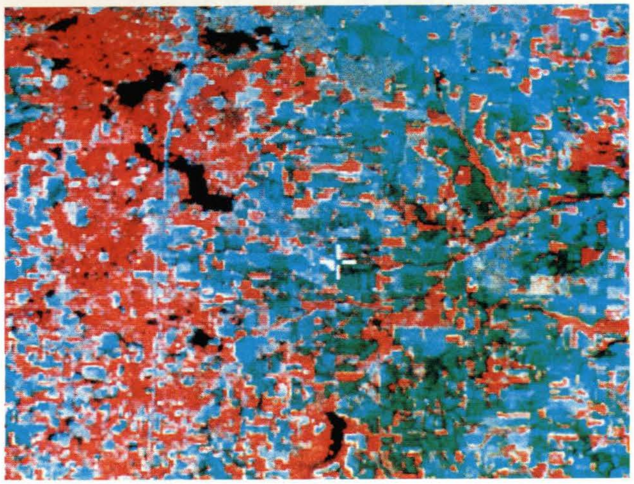
Orange - low building densities

Yellow - very low building densities

Various greens - trees, grass, hayfields

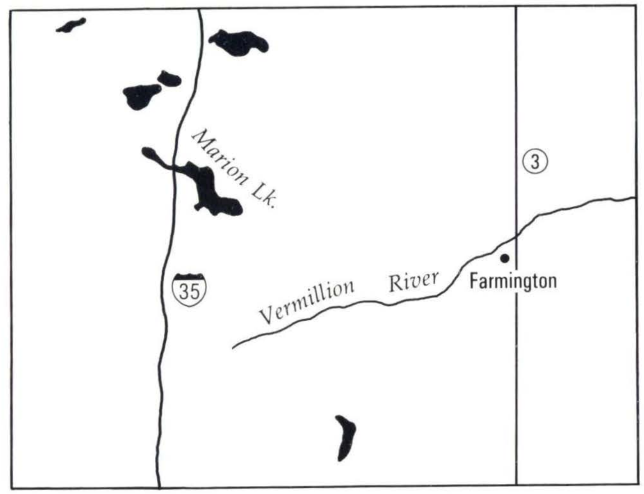
Dark brown - corn and soybean rowcrops

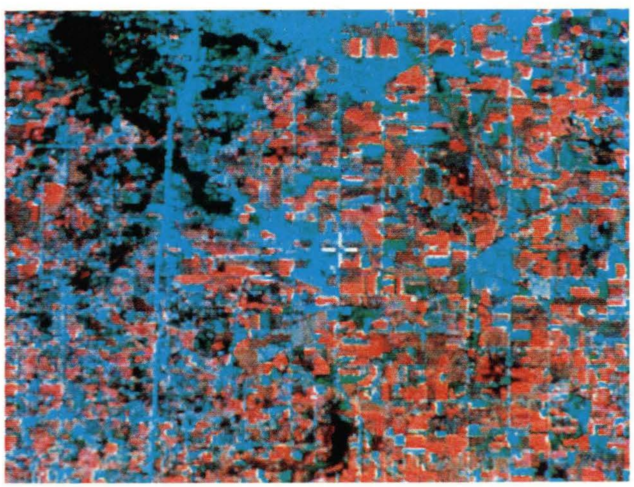
Plate 7-4



June
Image

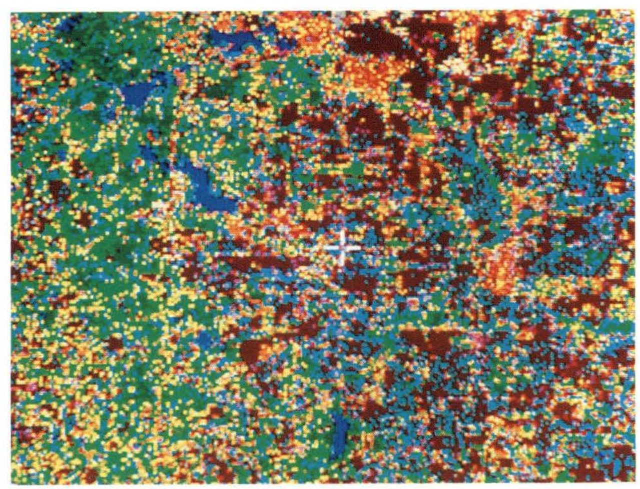
A





August
Image

B



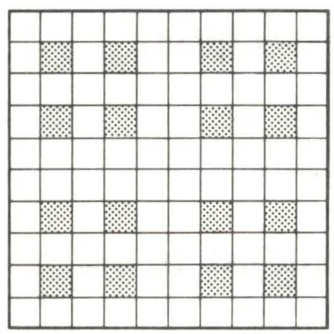
August
Classification

C

The soils and land cover data sets used in this study are not relational as they currently exist, even though they both were collected at kilometer resolution and use the same UTM coordinate base. The trouble lies in the different resolutions used to collect data in order to describe the characteristics of each kilometer. The soils data was collected using a sampling density of 16 points per square kilometer, whereas the land cover data has an apparent density of 100 pixels per square kilometer (see C. Gersmehl 1987).

In theory, the correction to this problem is simple: resample the land cover data at a density of 16 samples per km^2 using the pixels that match the soil sampling points. Figure 10 illustrates this approach.

FIGURE 10: LANDSAT PIXELS (100 BY 100 METERS IN SIZE) FOR LAND COVER CLASSIFICATION THAT CORRESPOND TO SOIL SURVEY SAMPLING POINTS WITHIN A SQUARE-KILOMETER CELL



A new inventory is then produced based on a point-counting of the *combination* of each land cover category and soil hydrologic group. Table 5 presents such a two-way frequency distribution for the Elm Creek basin on June 2nd and August 21st respectively; Table 6 presents the same type of data for the Vermillion River basin.

A major advantage of this relational approach, which uses the GIS to merge the two data sets, is in resolving differences in how certain categories are recorded during the sampling process. For instance, both the soil and land cover classifications have "water" as a category. These classifications should ideally overlap in their identification of water bodies. When you examine our GIS, however, very few samples actually match on their identification of water. Our solution is to call a place "water" if *either* classification identifies it as such. *We believe that it is better to err on the side of identifying too much surface water than to underestimate its extent.* In Minnesota, lakes are associated with hydrologic "sinks" -- water is held in the depression until it infiltrates into the groundwater system. A failure to identify these lakes may lead to a significant underestimate of the interaction between the surface and underground systems. Table 7 illustrates how the amount of water recorded for the two basins varies with the source of data used to identify its existence. The Vermillion River basin actually increased its allocation of water by having a number of soil samples identify water for places which were not water in the land-cover data.

TABLE 5: PERCENTAGES OF SAMPLE OBSERVATIONS WITH VARIOUS COMBINATIONS OF LAND COVER AND SOIL HYDROLOGIC GROUP: ELM CREEK BASIN FROM LANDSAT IMAGES*

JUNE 2, 1986

Land Cover Category		Soil Hydrologic Group			
		A	**B**	**C**	**D**
21	Extractive (gravel pits, dirt roads)	<1	<1	0	<1
30	Surface Water**	0	<1	<1	<1
41	Forested	0	2	<1	1
42	Grasslands	0	1	<1	<1
50	Wetlands	<1	4	1	4
61	Cover Crop	<1	13	4	8
63	Row Crop	0	12	3	6
71	Commercial/Industrial Bldg Density	0	1	<1	<1
72	High Building Density	0	2	1	2
73	Medium Building Density	<1	3	1	2
74	Low Building Density	0	4	1	3
75	Very Low Building Density	<1	6	1	2

AUGUST 21, 1986

Land Cover Category		Soil Hydrologic Group			
		A	**B**	**C**	**D**
21	Extractive (gravel pits, dirt roads)	0	<1	0	<1
30	Surface Water**	0	1	<1	<1
41	Forested	<1	4	1	1
42	Grasslands	0	3	1	3
50	Wetlands	0	4	1	3
61	Cover Crop	0	9	3	6
63	Row Crop	0	15	4	9
71	Commercial/Industrial Bldg. Density	<1	1	<1	<1
72	High Building Density	0	1	<1	1
73	Medium Building Density	<1	2	<1	1
74	Low Building Density	<1	6	2	2
75	Very Low Building Density	<1	4	1	2

*Out of 3712 sample points, 306 (8 percent) were unusable due to uncertain or undefined soil hydrologic group classification.

**Not included in this table are an additional 61 soil records (2 percent) that were recorded as "water" in the soil file, but classified as "non-water" on the land-cover file. These areas were assumed to be correctly identified as "water" by the soil survey.

TABLE 6: PERCENTAGES OF SAMPLE OBSERVATIONS WITH VARIOUS COMBINATIONS OF LAND COVER AND SOIL HYDROLOGIC GROUP: VERMILLION RIVER BASIN FROM LANDSAT IMAGE*

JUNE 2, 1986

Land Cover Category	Soil Hydrologic Group A	B	C	D
21 Extractive (gravel pits, dirt roads)	<1	<1	0	<1
30 Surface Water**	<1	<1	0	<1
41 Forested	<1	4	0	2
42 Grasslands	<1	1	0	1
50 Wetlands	1	9	0	5
61 Cover Crop	<1	12	<1	5
63 Row Crop	1	25	0	6
71 Commercial/Industrial Bldg. Density	<1	1	0	<1
72 High Building Density	<1	2	0	1
73 Medium Building Density	<1	4	0	1
74 Low Building Density	<1	5	0	1
75 Very Low Building Density	<1	5	0	1

AUGUST 21, 1986

Land Cover Category	Soil Hydrologic Group A	B	C	D
21 Extractive (gravel pits, dirt roads)	<1	<1	0	<1
30 Surface Water**	<1	1	0	<1
41 Forested	<1	5	0	2
42 Grasslands	2	4	0	2
50 Wetlands	1	6	<1	2
61 Cover Crop	<1	7	0	2
63 Row Crop	1	28	<1	11
71 Commercial/Industrial Bldg. Density	<1	2	0	<1
72 High Building Density	<1	2	0	<1
73 Medium Building Density	<1	3	0	0
74 Low Building Density	<1	8	0	1
75 Very Low Building Density	<1	3	0	1

*Out of 4958 sample points, 80 (2 percent) were unusable due to uncertain or undefined soil hydrologic group classification.

**Not included in this table are an additional 158 soil records (3 percent) that were recorded as "water" in the soil file, but classified as "non-water" on the land cover file. These areas were assumed to be correctly identified as "water" by the soil survey.

TABLE 7: OCCURRENCE OF "WATER" (BY PERCENT OF BASIN AREA) ACCORDING TO THE SOIL DATA, LAND-COVER CLASSIFICATION, AND THE COMBINED RELATIONAL SOIL AND LAND-COVER DATA SETS

		Land Cover		Relational Data Set	
	Soils	June	August	June	August
Elm Creek Basin	2	2	3	2	3
Vermillion River Basin	3	2	3	4	4

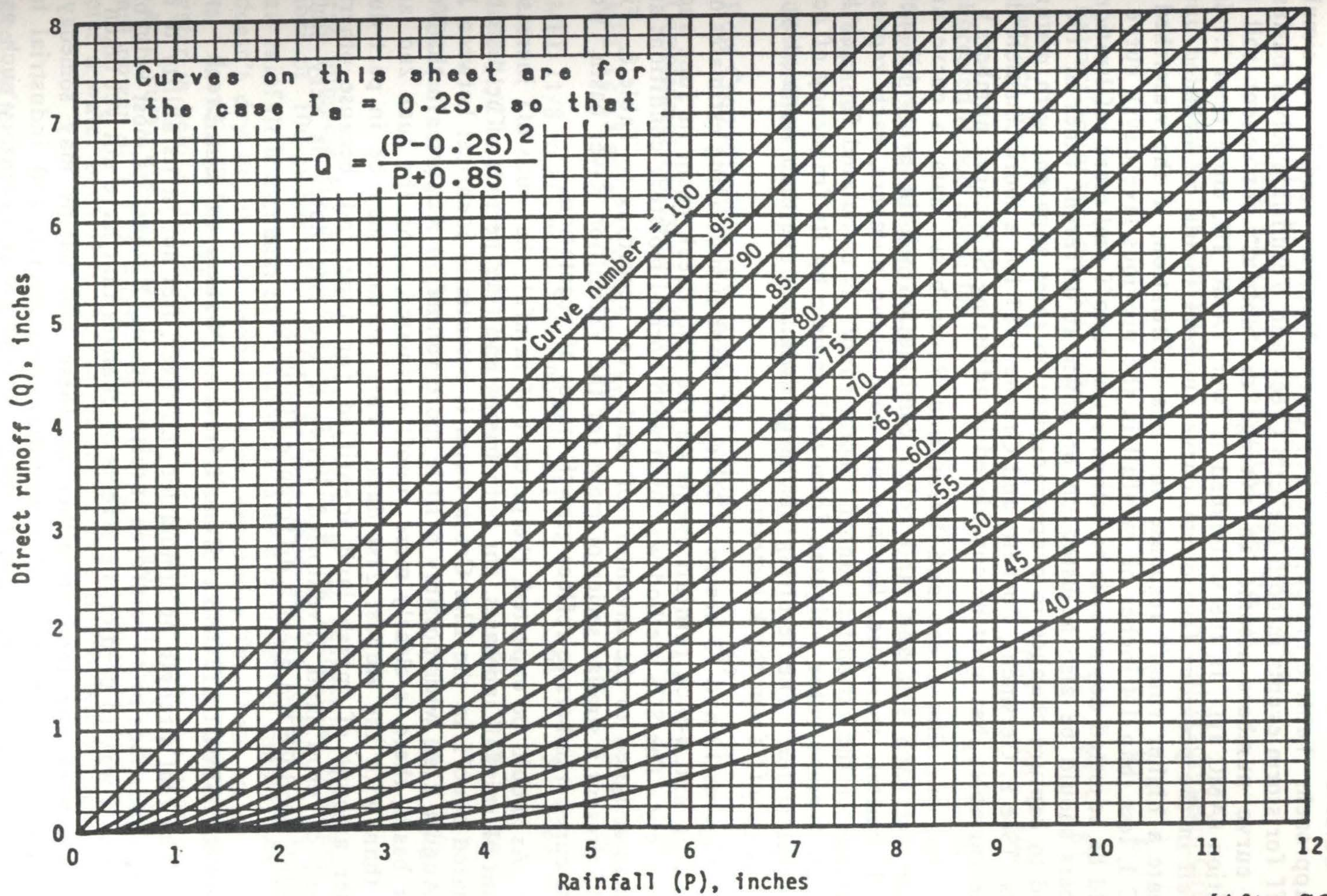
Three explanations exist for these disparities in relating the two data sets. **First**, the dates of the county soil surveys and the Landsat flights are not the same. These differences exist both for the year (e.g., 1959 for the Scott County survey, 1986 for the Landsat imagery) and for the season within the year (springtime photography for soil surveys and June and August satellite imagery for the land cover data). Extent of surface water varies continuously through the year, as a result of changes in antecedent moisture conditions and rainfall amounts. Soil surveys will tend to identify long-term wetlands, whereas the Landsat classification will capture short-term events. **Second**, a land-cover classification from Landsat imagery has the potential for several mis-classifications due to problems in identification of good spectral signatures. It is particularly troublesome to classify water during the summer months because of the varying amount of vegetative and biological growth present in the water bodies. This "green growth" affects passive reflectance values recorded in the infrared range of light and depends heavily on a number of factors, including the depth and size of the water body. And **third**, there are problems in the ability to register Landsat imagery from two time periods to the same coordinate base. Even though the same UTM-based coordinate system is used to gather ground control points for a rectification of a Landsat image or aerial photographs, the apparent locational precision is misleading. For the Elm Creek area, edges of water bodies were often shifted by one pixel to the north, south, east, or west between the the two time periods. This shift was inconsistent across the basin, indicating that the cause is not easy to identify and therefore a simple correction within the GIS may not be possible (C. Gersmehl 1987). The simplest solution, therefore, is to leave the two images as they are (make no attempt to correct the registration) and assume that they are correct for an inventory of the land-cover categories within the kilometer.

We can test the validity of that assumption only in the context of the use for which we are gathering the data. We therefore will turn to a discussion of the simulation methods that will be used with these data.

SIMULATION OF RUNOFF

A simple (but not necessarily the best) way to estimate runoff for a basin is to use the Soil Conservation "Curve Number" approach for ungauged watersheds (Figure 11). The SCS method uses the idea of *hydrologic soil-cover complexes* (combinations of hydrologic soil groups with land cover and treatment classes) for each location (Soil Conservation Service 1972, 1979, and 1986; Kent 1973). Each soil-land cover complex is empirically related to a *curve number* (CN) between 25 and 100. (Figure 11 illustrate runoff from the curve numbers between 40 and 100.)

DIRECT RUNOFF ESTIMATED
USING THE SOIL CONSERVATION SERVICE "CURVE NUMBER" APPROACH



[After SCS 1986]

FIGURE 11

The higher the curve number, the greater the amount of runoff that is likely to occur for a given rainfall event.

Calculation of direct runoff from precipitation is straightforward using the SCS approach. Table 8 presents the SCS Curve Numbers used in our simulation of runoff for storm events in both the Elm Creek and Vermillion River basins. All of these curve numbers assume an average or "normal" antecedent soil moisture condition (AMC II). If the watershed is "wet" (AMC III, more than 2.1 inches of rainfall in the previous five days), curve numbers should be adjusted upwards to estimate a higher runoff response and less infiltration; if the watershed is "dry" (AMC I, less than 1.4 inches of rainfall in the previous five days), curve numbers should be adjusted downwards to estimate less runoff and more infiltration. Curve numbers should be also adjusted up or down at different times of the growing period to capture changes in evapotranspiration rates apparent with maturation of plants. (See P. Gersmehl, Corbett, and Greene 1987 for a more complete discussion of the variables that must be considered in adjusting SCS curve numbers.)

For our Elm Creek and Vermillion River basin data, a curve number is derived for each soil and land cover combination within every kilometer area. Each kilometer, therefore, has one or more curve numbers, each representing a percentage of the estimated runoff. Total runoff is calculated for each kilometer using an average value weighted from each sample by its frequency of occurrence in the kilometer. The resulting discharge estimates are then summed to estimate the total runoff from the kilometer.

Figures 12 and 13 present a series of maps summarizing estimates of runoff for 1, 2, 3, and 4-inch rainfall events on the Elm Creek basin in June and August respectively. The SCS method was used with normal moisture conditions and curve numbers assigned to the inventory of land cover according to Table 8. Figures 14 and 15 present a similar series of maps for the Vermillion River basin. We did not adjust curve numbers to reflect growth stages of vegetation.

As should be expected, the percent of rainfall that runs off increases as the amount of rainfall increases, for both basins and time periods. The distribution of estimated runoff in the Elm Creek basin is essentially identical between June 2nd and August 21st, with possibly a little more runoff in August. In the Vermillion River basin, there is clearly more runoff estimated using the June 2nd land cover data than with the August 21st data. This increase may be in part due to the greater amount of marsh land identified for June 2nd, because otherwise the amount of crop land inventoried between the two dates is very similar and shouldn't produce too great a difference (Table 6).

"WHAT-IF" SIMULATIONS -- Runoff from an "Urbanized" Elm Creek

The real strength of simulation, however, lies not in its ability to duplicate existing conditions or to reinforce intuitive ideas. Rather, its most valuable role is to serve as a way to predict the effects of changes that have not yet taken place.

Much of the land in either of these two basins may someday become urbanized, with a mixture of residential, commercial, and industrial land uses replacing the current agricultural uses. A land cover inventory such as the one presented in this paper, when matched with a simulation model to estimate runoff, is very useful in testing the effects of various growth scenarios before they even take place. The results of these studies can then be used to modify current policies or plans to direct this new growth in a fashion that meet specific goals. These goals might be, for instance, to limit the total increase in runoff to a specific

TABLE 8: CURVE NUMBERS USED WITH THE LANDSAT LAND-COVER DATA

Land Cover Category	Soil Hydrologic Group A	B	C	D
21 Extractive (gravel pit, dirt road)	72	82	87	89
30 Surface Water*	100	100	100	100
41 Forested (Good condition)	25	55	70	77
42 Grasslands (good condition)	39	61	74	80
50 Wetlands	85	85	85	85
61 Cover Crop (Straight row, good condition)	58	72	81	85
63 Row Crop (Straight row, good condition)	67	78	85	89
71 Commercial/Industrial (85% impervious)	89	92	94	95
72 HBD (<1/8 acre, 65% impervious)	77	85	90	92
73 MBD (1/8 to 1/2 acre, 35% impervious)	61	75	83	87
74 LBD (1/2 to 1 acre, 25% impervious)	54	70	80	85
75 VBD (>1 acre, 20% impervious)	51	68	79	84

*Curve numbers for surface water present a big theoretical problem. A curve number of 100 (what SCS assigns to this land cover) implies "maximium runoff" -- all rainfall is assumed to be automatically available for runoff. As long as a lake doesn't dry up, this is probably true, especially for those water bodies with outlets. In Minnesota, however, lakes are also associated with hydrologic "sinks" -- water is held in the depression until it infiltrates into the groundwater system. A zero curve-number may be more appropriate. Here is one area where the relational power of a GIS might come in handy -- a network file will tell us how connected a water body is to the rest of the surface hydrologic system.

(Source: After Soil Conservation Service 1972 and 1986; Young et al. 1985)

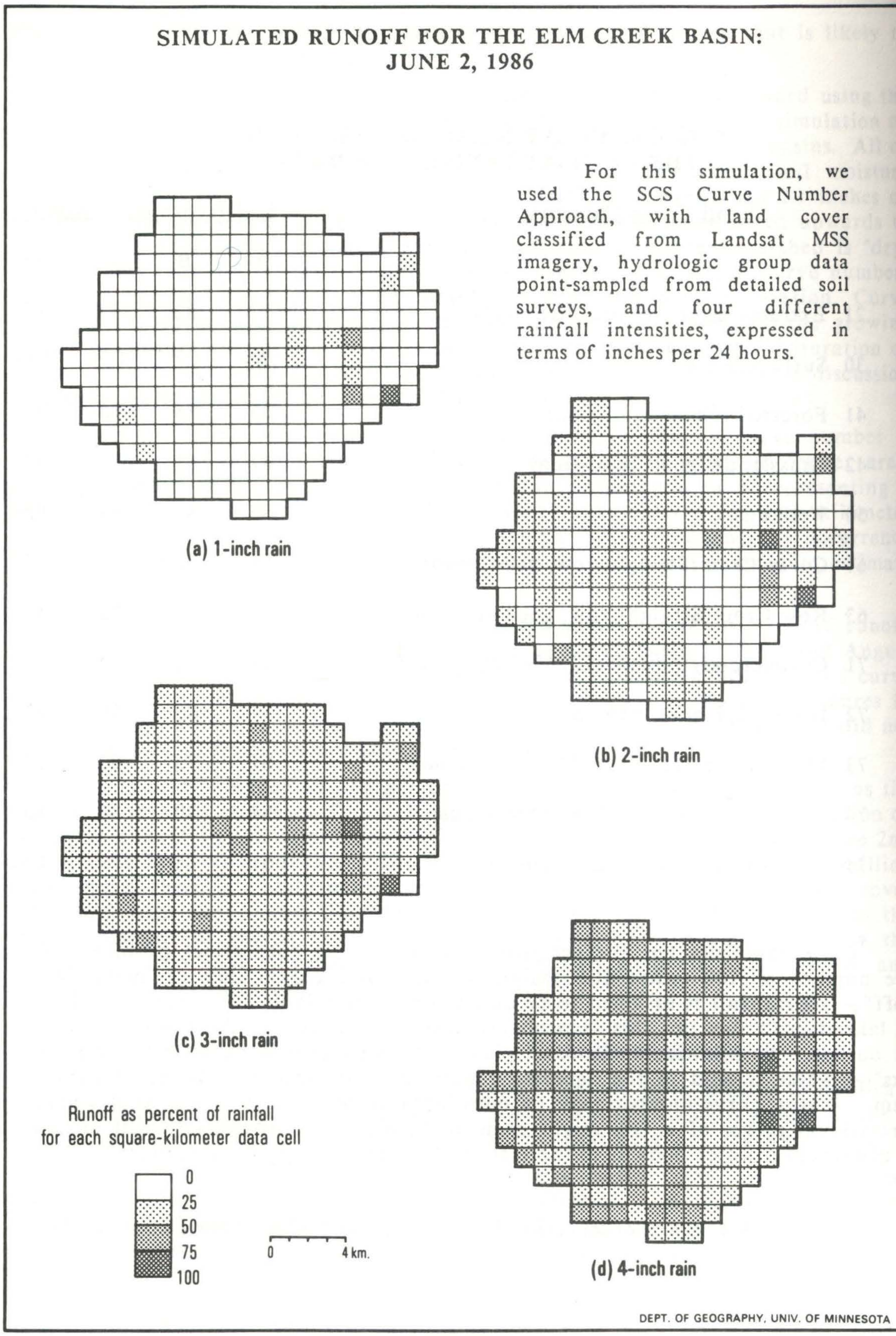

SIMULATED RUNOFF FOR THE ELM CREEK BASIN:
JUNE 2, 1986
For this simulation, we used the SCS Curve Number Approach, with land cover classified from Landsat MSS imagery, hydrologic group data point-sampled from detailed soil surveys, and four different rainfall intensities, expressed in terms of inches per 24 hours.
(a) 1-inch rain
(b) 2-inch rain
(c) 3-inch rain
(d) 4-inch rain
Runoff as percent of rainfall
for each square-kilometer data cell
0
25
50
75
100
0
4 km.
DEPT. OF GEOGRAPHY, UNIV. OF MINNESOTA

FIGURE 12

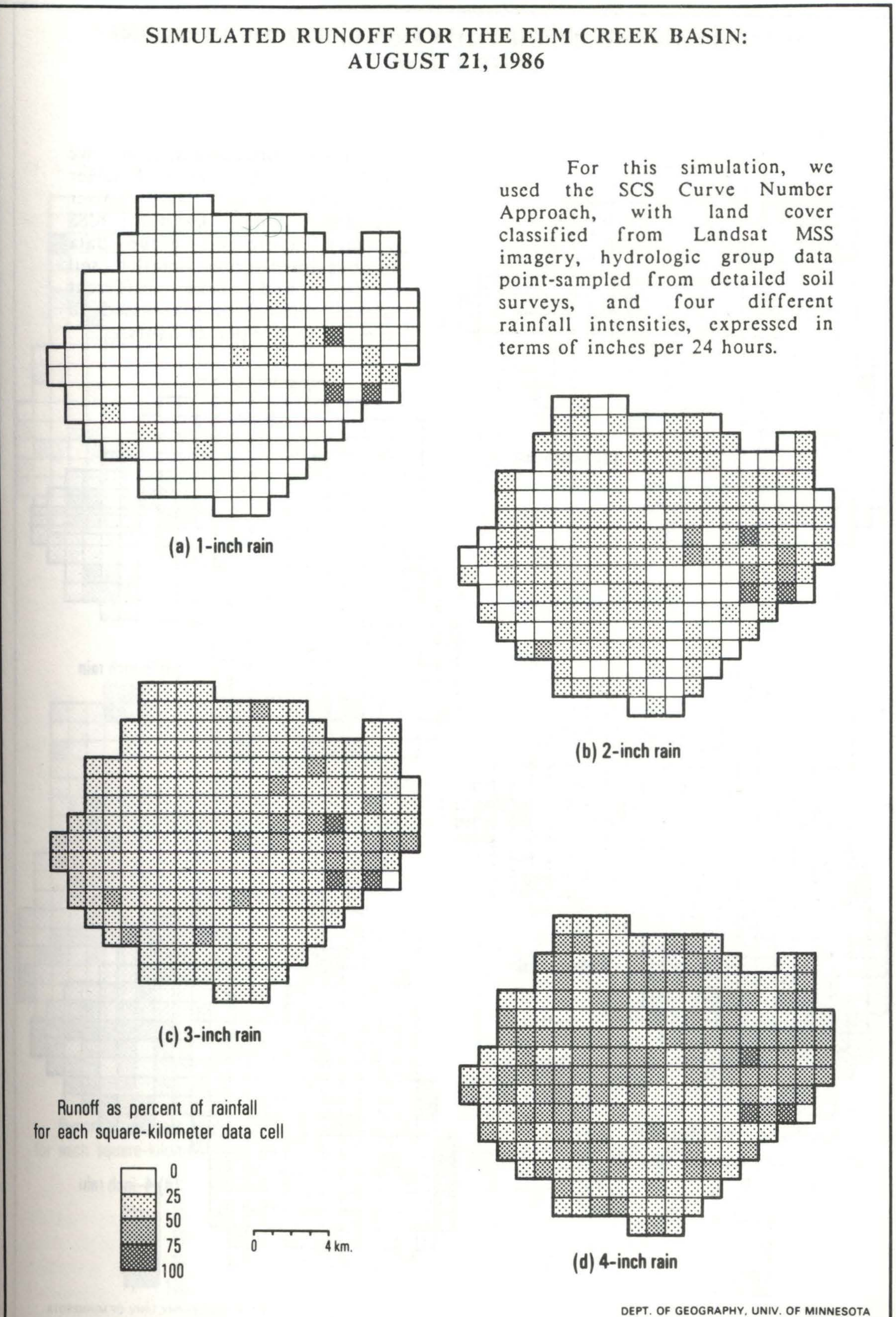
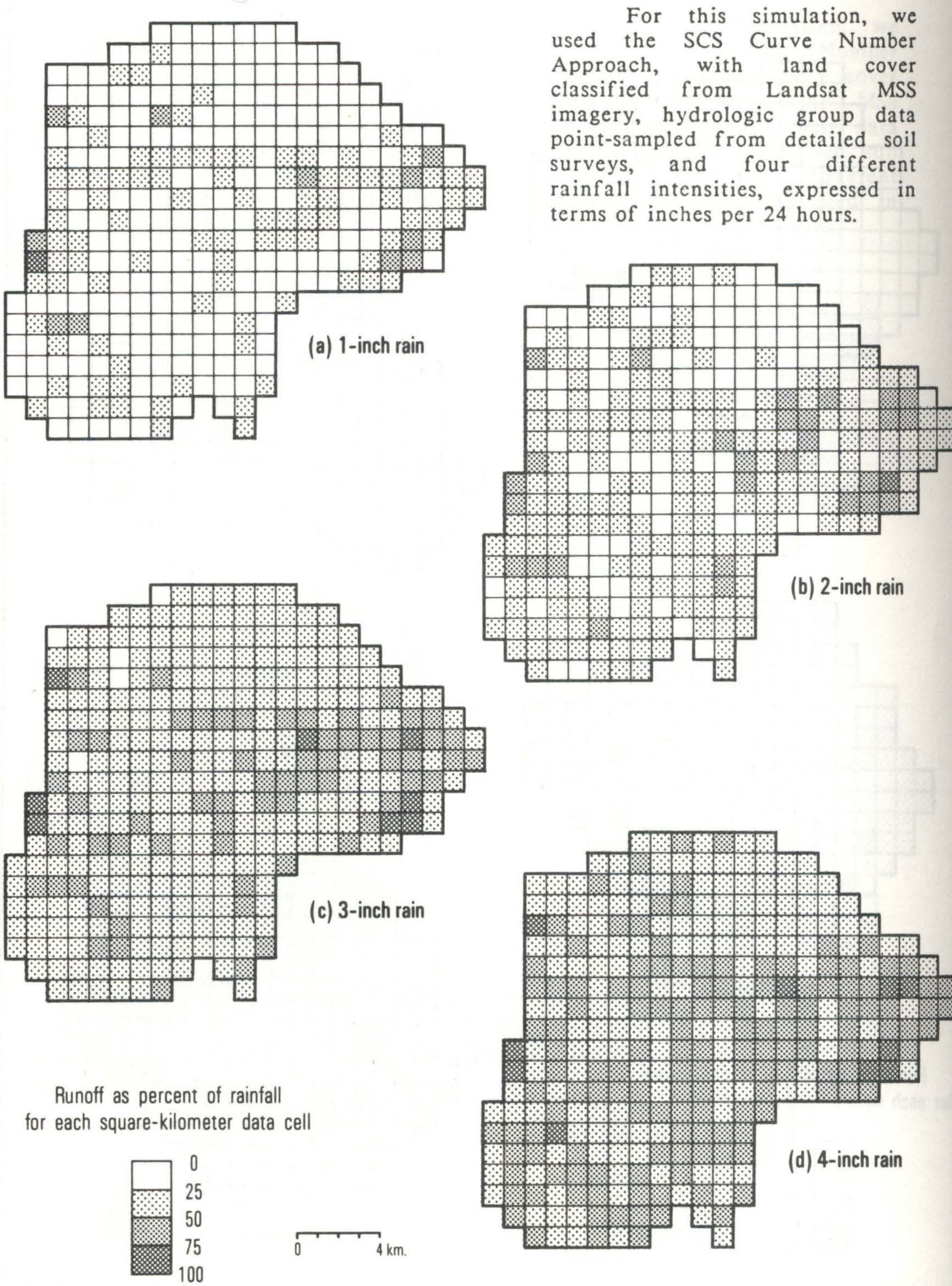

SIMULATED RUNOFF FOR THE ELM CREEK BASIN:
AUGUST 21, 1986
For this simulation, we used the SCS Curve Number Approach, with land cover classified from Landsat MSS imagery, hydrologic group data point-sampled from detailed soil surveys, and four different rainfall intensities, expressed in terms of inches per 24 hours.
(a) 1-inch rain
(b) 2-inch rain
(c) 3-inch rain
(d) 4-inch rain
Runoff as percent of rainfall
for each square-kilometer data cell
0
25
50
75
100
0
4 km.
DEPT. OF GEOGRAPHY, UNIV. OF MINNESOTA

FIGURE 13

**SIMULATED RUNOFF FOR THE VERMILLION RIVER BASIN:
JUNE 2, 1986**

For this simulation, we used the SCS Curve Number Approach, with land cover classified from Landsat MSS imagery, hydrologic group data point-sampled from detailed soil surveys, and four different rainfall intensities, expressed in terms of inches per 24 hours.



DEPT. OF GEOGRAPHY, UNIV. OF MINNESOTA

FIGURE 14

SIMULATED RUNOFF FOR THE VERMILLION RIVER BASIN: AUGUST 21, 1986

For this simulation, we used the SCS Curve Number Approach, with land cover classified from Landsat MSS imagery, hydrologic group data point-sampled from detailed soil surveys, and four different rainfall intensities, expressed in terms of inches per 24 hours.

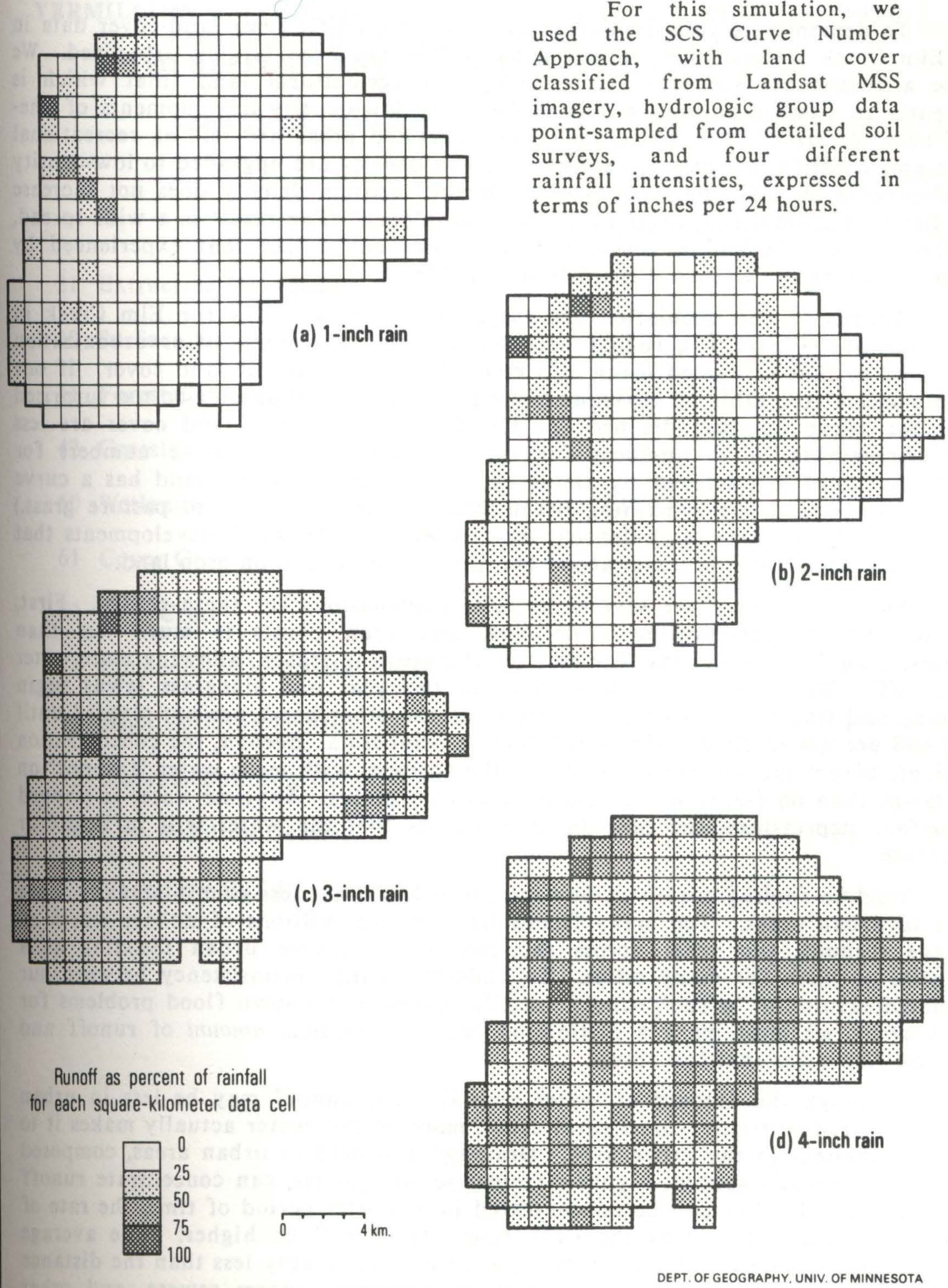


FIGURE 15

tolerable level, or to minimize the reduction of water available for infiltration in order to insure a minimum volume of groundwater for domestic supplies. Whatever the goal, it is certainly cheaper to spend some money now for computer time in order to test these scenarios than to spend a lot of money later to correct some mistakes.

To demonstrate this simulation process, we modified our land-cover data in the Elm Creek watershed to reflect a basin that has been largely urbanized. We chose a reasonable scenario for growth: all agricultural land (that which is currently in pasture or cropped) is replaced by low-density developments of one-half acre lot sizes. Wet lands, lakes, and forested areas are left as recreational areas and wildlife refugees. Areas of very low density are upgraded to low-density developments by infilling. Areas of high and medium density does not increase and little additional industrial property is developed. The result is a wide-spread, continuous distribution of residential properties not unlike that experienced by many of our nation's cities (Figure 16 and Table 9).

Results of the simulation are surprising. As the maps for Elm Creek in June indicate (Figure 17), runoff actually appears to *decrease* after urbanization compared to that estimated using SCS methods for the current land cover. If one looks closer at the table of curve numbers (Table 8), one shouldn't be too surprised since the curve numbers defined for low-density residential land cover are less than comparable curve numbers for land in row crops. (Curve numbers for residential land are derived by assuming that the impermeable land has a curve number of 98 and the remaining permeable land is equivalent to pasture grass.) It's not until you have high-density residential or commercial developments that the amount of runoff increases above the amounts produced on crop land.

Physically, there are several possible explanations for these results. **First**, the amount of transpiration from corn and bean plants is much less than transpiration from lawn grass for most of the growing season, allowing more water to runoff. We must realize how early in the year our bluegrass lawns begin growing and transpiring water; most crops, on the other hand, are not planted until May and are not really growing until June in Minnesota. **Second**, the concentration of plants affect the capture of water on the surface, with more water "trapped" on our lawns than on fields of sparsely-planted corn stalks. Anytime water is trapped in surface depressions, the probability increases that the water will infiltrate or evaporate.

We'd be fooling ourselves, however, to believe that less runoff actually takes place in urban areas. Too many floods have caused millions of dollars worth of damage through the years across the country to ignore urban flooding as a potential problem for many cities. To understand this inconsistency between our simulated results of "urban" runoff for Elm Creek and known flood problems for urban areas, we must make a distinction between the *total amount* of runoff and the *peak flow rates* that can result.

Although the amount of water available for runoff may be less in urban areas compared to similar-sized rural areas, more of this water actually makes it to a stream channel in the urban area. Drainage networks in urban areas, composed primarily of closely-spaced, highly-efficient sewer systems, can concentrate runoff much quicker. If flow can be concentrated in a shorter period of time, the rate of flow in the channel at that the peak time period will be higher. The average distance from a homeowner's lawn to a gutter is considerably less than the distance from a farmer's field to the nearest ditch. Gutters, storm sewers, and other channels in urban areas are enclosed and made of concrete and other impervious

TABLE 9: OCCURRENCE OF LAND-COVER TYPES IN THE ELM CREEK AND VERMILLION RIVER BASINS AFTER SIMULATED URBAN GROWTH: CONVERSION OF ALL AGRICULTURAL LAND TO DEVELOPMENTS OF LOW BUILDING DENSITY

Land Cover Category	Percentage of Basin Area			
	Elm Creek		Vermillion River	
	June	**August**	**June**	**August**
21 Extractive (gravel pits, dirt roads)	1	1	1	1
30 Surface Water	2	3	2	3
41 Forested	3	7	6	8
42 Grasslands	1	5	1	3
50 Wetlands	11	9	16	9
61 Cover Crop	-	-	-	-
63 Row Crop	-	-	-	-
71 Commercial/Industrial Bldg. Density	2	2	1	2
72 High Building Density	4	3	3	3
73 Medium Building Density	5	4	5	4
74 Low Building Density	70	67	66	68
75 Very Low Building Density	-	-	-	-

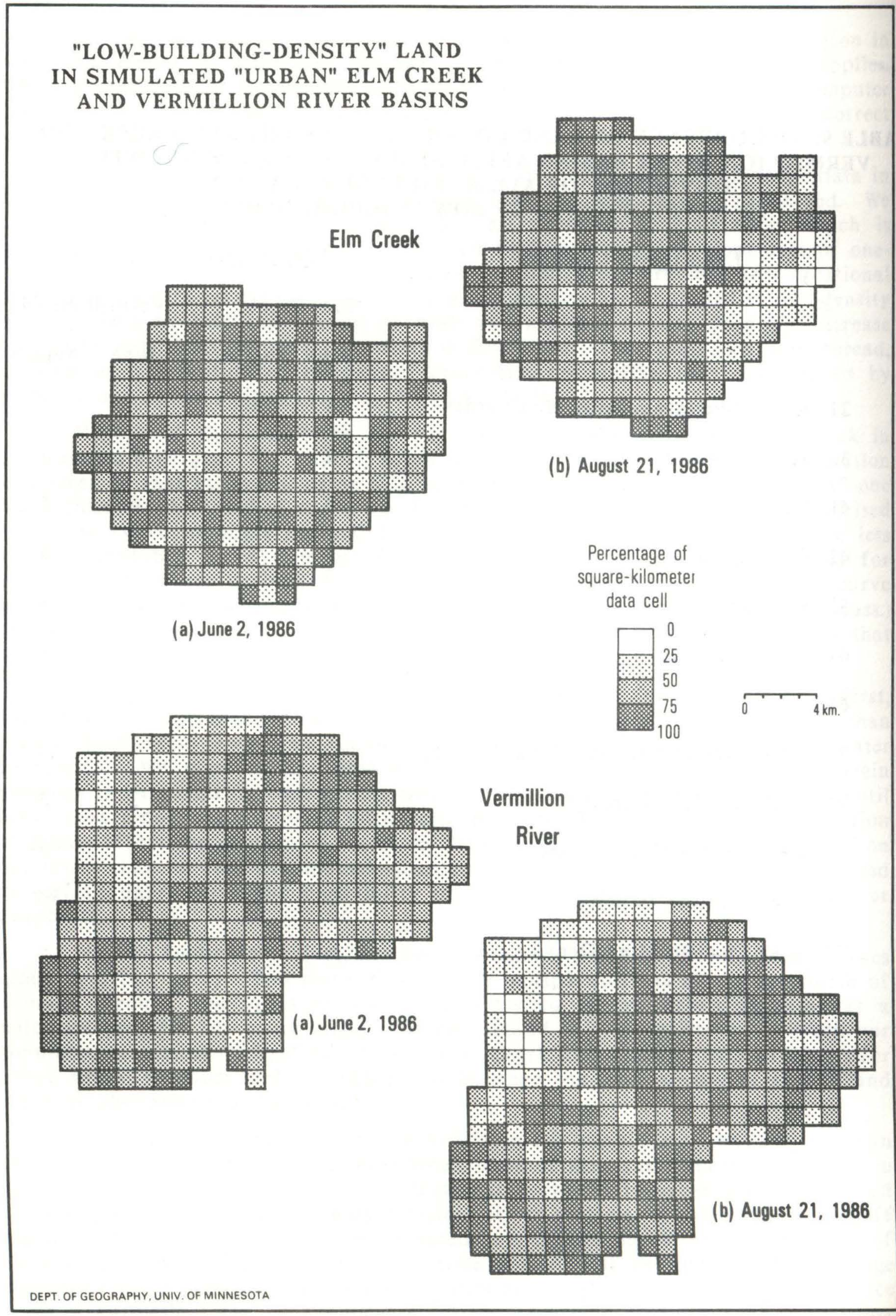
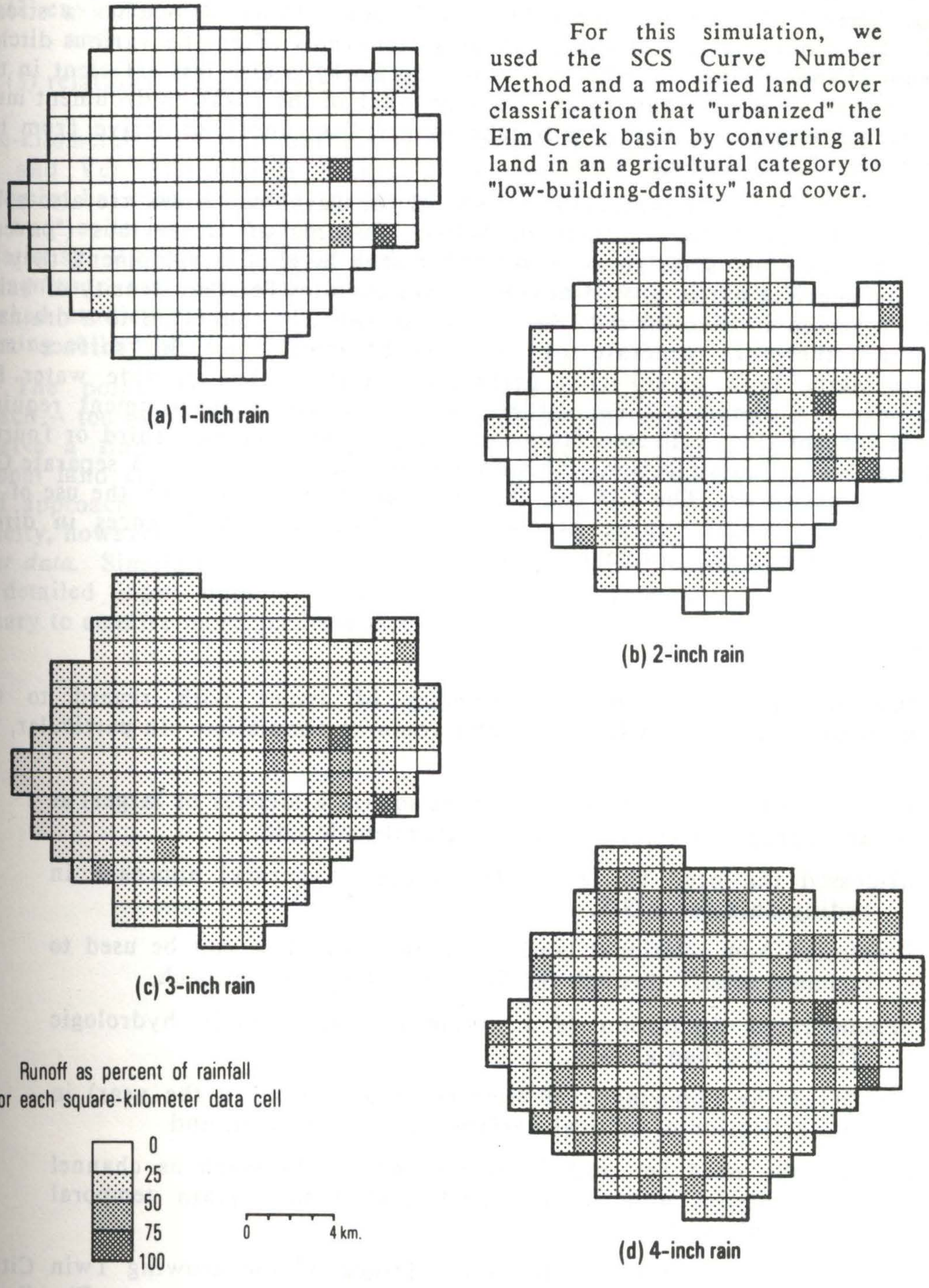

"LOW-BUILDING-DENSITY" LAND
IN SIMULATED "URBAN" ELM CREEK
AND VERMILLION RIVER BASINS
Elm Creek
(a) June 2, 1986
(b) August 21, 1986
Percentage of
square-kilometer
data cell
0
25
50
75
100
0
4 km.
Vermillion
River
(a) June 2, 1986
(b) August 21, 1986
DEPT. OF GEOGRAPHY, UNIV. OF MINNESOTA

FIGURE 16

SIMULATED RUNOFF FOR THE "URBAN" ELM CREEK BASIN: JUNE 2, 1986

For this simulation, we used the SCS Curve Number Method and a modified land cover classification that "urbanized" the Elm Creek basin by converting all land in an agricultural category to "low-building-density" land cover.



DEPT. OF GEOGRAPHY, UNIV. OF MINNESOTA

FIGURE 17

materials; almost all urban runoff will find its way to a stream, lake, or other water body. Baseflow (constant, year-round flow in a channel) is minimal or does not exist in an urban sewer system, largely as a result of the impervious nature of these channels.

Country ditches, on the other hand, are open, earth-lined structures, with high infiltration rates through their bottom and sides. Most of the runoff in rural areas has plenty of time to evaporate or infiltrate before it reaches a stream channel. The "connectivity" of the drainage system (how often the various ditches and stream channels come together to form a network) is also less efficient in the rural areas than in most urban areas. Each channel in the rural environment must drain a larger area, and thus it will take longer for runoff to arrive from the farthest edges of the drainage basin.

The curve numbers presented in Table 8 for developed areas are also a bit misleading. Although they assume particular amounts of impervious ("paved") surface and open land (some form of grass) for each level of development, they do not describe specific patterns of drainage within each lot for that density of built-up land. A paved surface (or a building) may be directly connected to a drainage system (and therefore generate direct surface runoff), or the surface may contribute water to a neighboring pervious area (and thus provide water for infiltration). To simulate these differences in an urban environment requires additional information, collected from field observations, from a third or fourth-level of land-cover classification. This data could also be stored in a separate GIS file of sewer networks. The GIS can then be used, possibly through the use of an expert-system, to modify curve numbers to reflect these differences in direct contribution to the urban drainage system.

SUMMARY

This report has presented a number of issues with regard to the development of data sets in a GIS for water resource simulation. In particular, we have:

- focused on the use of coarse cells (one square kilometer or larger) as an appropriate resolution in a regional-scale GIS,
- discussed the need for a point-counting (inventory) approach in coding data for a GIS,
- illustrated how relational land cover and soils data can be used to estimate runoff through the SCS curve number approach,
- demonstrated simulation as a viable tool to identify hydrologic effects of urban growth,
- identified the importance of time (such as season in the year) in data collection of critical variables like land cover, and
- realized the need for additional relational data (such as channel characteristics and routing coefficients) to explain temporal patterns of runoff.

Two medium-sized watersheds on the fringe of the growing Twin Cities Metropolitan Area served as a "test-bed" for new data file structures. Elm Creek and Vermillion River watersheds (of Hennepin and Dakota Counties respectively) were appropriate for the task. The purpose of this particular report was not to define file structures (that has been left to our companion report, Brown and Gersmehl 1987), but illustrate their usefulness in an urbanizing area.

Three major conclusions can be drawn from this study:

(1) useful hydrologic analysis is possible with a simple GIS and simulation method,

(2) environmental data should be coded by the point-counting method, and

(3) relational data sets are a must.

Square-kilometer cells were sufficient for analysis of surface runoff in the Elm Creek and Vermillion River basins. Our choice of kilometer resolution is a compromise between the need for *detail* (in identifying the variety of land-cover types and soils in a basin) and *simplicity* (in data collection and runoff calculations). Runoff calculated and mapped as a percentage of rainfall provides the magnitude and direction of runoff response from changes in land cover. Watershed studies cannot be any more accurate without the costly and time-consuming collection of field observations.

The point-counting method is necessary with the SCS curve-number approach -- too many different cover-soil combinations exist in a square kilometer to derive a single dominant combination. The job cannot be done without relational land cover and soil data. The two needs go hand-in-hand. The SCS runoff approach provides plenty of useful information for policy decisions. Its simplicity, however, must be handled with care. *Do not put blind trust in its results or your data.* Simulation results cannot be displayed any more accurately than the least detailed input data. If more detail is required, field observations are necessary to acquire more accurate data.

APPENDIX

CREATING A POINT-COUNT INVENTORY IN EPPL7

Kevin L. Anderson

Introduction

EPPL (Environmental Planning Programming Language), developed by the Minnesota State Planning Agency's Planning Information Center (PIC), comes in two versions: EPPL6, the "big brother" program on the Prime computer, and EPPL7, the "junior" version for 16-bit microprocessors. Neither version currently implements point-counting or other cell-based inventory logic as a normal option. Inventories in a GIS are files that record the proportional occurrence of phenomena in an area (*point-counting* approach) as opposed to assigning a single category to each area (*area-tagging* approach). EPPL, like most GIS packages on the market, follow the traditional approach of tagging each area with a single category for each variable (P. Gersmehl, Brown, and Anderson 1987).

Inventories are possible in EPPL. They can be created with just a few simple commands. What follows is an outline of the sequence of commands that can be used to implement a point-counting file structure in EPPL7.

File Structures

Understanding how an inventory is developed in EPPL7 requires some background on the file structure of the current package. The EPPL programs use a *flat file structure* for their GIS. A flat file, like spreadsheets and other two-dimensional objects including maps, have no *depth* to them. A variable in the GIS (which is functionally equivalent to a single-theme printed map), has all of its information stored in one layer of a file. Each cell, polygon, or unit area in the GIS has only a single record in the file, which describes it in terms of the category in which it appears to "belong" (i.e. the category that occupies the most area or "dominates" the cell in some other way). Thus, a land-use file might consist of a mass of letters to designate the dominate land use in each cell (e.g. "cultivated", "forested", etc.).

Inventories, on the other hand, record the proportional importance of every category in each cell of the GIS. This approach normally requires a file structure with multiple layers (or "depth"), one layer for each possible category of the variable. Each layer contains the occurrence, in percent, of that category in each data cell. The percentage values are calculated on the basis of either the total area of the cell (if the areal extent of each category was measured) or the total number of sample points in the cell (if a point-counting technique was used to analyze the phenomena).

In a flat-file system such as EPPL, several files must be used in concert to represent a full inventory of a single variable. As many files must exist as there are categories in the variable (e.g. nine categories is the "universe" for the state-wide land use data). The number of categories in a file may not be the same, however, as the highest number assigned to a category. For example, the land-cover classification in Table A-1 has 14 categories, with the highest category labeled as 75; the land-cover inventory will have 14 files, one for each category. On PC-DOS/MS-DOS computers, such as the IBM AT, EPPL7 has been implemented to include the possibility of up to 256 different categories for any single variable. On the Prime Computer, up to 32,000 or more categories are possible with EPPL6. In most cases, however, (with the possible exception of elevation data) very few categories are actually needed for a variable.

TABLE A-1: LAND-COVER CATEGORIES USED IN THE EXAMPLE INVENTORY

10 Hard Surface, Undifferentiated	61 Cover Crop
	62 Small Grains
21 Gravel pits, Extractive	63 Row Crops
30 Surface Water, Undifferentiated	71 Commercial/Industrial
	72 High Building Density
41 Forested, Woods, Trees	73 Medium Building Density
42 Grasslands	74 Low Building Density
	75 Very Low Building Density
50 Wetlands, Undifferentiated	

The only real limitation to implementing an inventory system with EPPL7 is in PC-DOS/MS-DOS file-naming restrictions. DOS allows only eight (8) characters for each file name. In an inventory system, one or two characters must be set aside to represent the category number stored in this file. That leaves only six significant characters to describe the variable itself (e.g. PINELU01 to represent category 1 of the Pine County land-use file). Three or four letters may be sufficient for Minnesota county names, but how are we going to handle townships or cities? Creative use of DOS sub-directories (similar to how it is currently done on the Prime) could alleviate some of this problem (at the expense of further fragmenting the logical structure of the hard disk), but it certainly won't solve it entirely. All files for a particular inventory would then be assigned their own DOS sub-directory (e.g. \MLMIS40\PINE\LUSE).

EPPL7 Command Sequence

Inventories can be created in EPPL7 (and EPPL6 for that matter) following four basic steps. Each step, however, requires execution of several functions. Five EPPL7 commands are used repeatedly in these steps: **reclass, rescale, jumping window, evaluate,** and **quit**. The example that follows contains sample output from EPPL7, plus typical responses to requests for parameters. I do not intend to describe each command in detail; that is the function of the EPPL7 user's manual, published by the Planning Information Center, Minnesota State Planning Agency. I intend only to provide the basic sequence of commands that can be adapted by users for their particular needs. All user responses are underlined, and *<cr>* means hit the carriage-return key for a blank response (empty line).

Step One: Prepare the Inventory Base File

An *inventory base file* is a file that records, for each data cell, the number of observation points or areas that are used to calculate the point-count inventory. Every variable in a given study area has a base file. The base file is created at the resolution of the final inventory, not that of the original data inventory. For example, consider a point-counting inventory with an output resolution of one kilometer and an input of 100 data points in each kilometer. The base file will have one kilometer-sized cell for every 100 cells in the input-data file. These 100 data cells are arranged in a 10 by 10 matrix.

The base file maintains a count of all *non-offsite* values in each cell. The **reclass** command is used to create a temporary version of the input-data file with all offsite cells recoded to 0 and other categories set to 1. The **jumping window** is then used to *count* those samples that are valid data. The **rescale** command is used after the count to reduce the cell resolution to inventory scale.

In the example that follows, we will create the base file ("eabase") from the land-cover data ("ealcov") for the Elm Creek basin in August. The land-cover file contains 100 samples (arranged 10 by 10) for an inventory at kilometer resolution. The **align** and **mosaic** commands have been previously used to align the UTM coordinates of the upper-left corner of the GIS land-cover file (cell 1,1) with a kilometer boundary (e.g. 445,000 meters east, 5,018,000 meters north).

```
EPPL7, beta release
Copyright (C) - September, 1986
Minnesota State Planning Agency

Command: reclass
Number of input files: 1
Old file: ealcov
New file: temp0
Description: *
Enter one reclass statement per line. Use empty line to end.
* 0 = 0 255
* 1 = 1:254
* <cr>
Beginning reclass...

Command: jumping
Old file: temp0
New file: temp1
Description: *
Horizontal window size: 10
Vertical window size: 10
Option: 1
Scale factor for result: 1

Command: rescale
Old file: temp1
File has dimensions: 411 rows & 351 columns
Enter resampling option, N)earestneighbor or B)ilinear: n
Column(x) scale factor: 0.1
Row(y) scale factor: 0.1
Output has dimensions of 41 rows & 35 columns
New file: eabase
Description: Elm August Base Sample Size
```

Step Two: Derive Inventory for the First Category

Counting the occurrences of a single category in a specified area requires three EPPL7 commands. We use the **evaluate** command to highlight each data point coded as that category. These points are set to 1 while points in all other categories are set to 0. The **jumping window** command counts the occurrence within the kilometer cell (or some other resolution used in the inventory) and the **rescale** command reduces the count to a single value through a change in scale. We calculate the percent occurrence of the category with the **evaluate** command by dividing its count by the sample base. This occurrence is stored as a value from 0 to 100 in the inventory file for this category. (An offsite value of 255 is used for easy identification during the mapping phase.) The file name and description should record the category name and/or number. (You may find, as I did very quickly, that the 32 characters defined for an EPPL7 file description are not enough to make an adequate record of what is stored in the file.)

In the example that follows, we will continue with the same data set for which the sample base file was just created. Values in this base file become the divisor in the calculation of percent occurrence. The category values whose inventory is being created is category 30 ("surface water, undifferentiated"). Again, we have 100 data samples for each kilometer.

```
Command: evaluate
Number of old files:  1
Old file:  ealcov
Number of new files:  1
New file 1
 Offsite value:  255
 File name:  temp0
Description:  *
Enter evaluate statements.  Use empty line to end.
*  if old1 = 30 then
*    new1 := 1
*  else
*    new1 := 0
*  end;
* <cr>
Evaluating.....

Command: jumping
Old file:  temp0
New file:  temp1
Description:  *
Horizontal window size:  10
Vertical window size:  10
Option:  1
Scale factor for result:  1

Command: rescale
Old file:  temp1
File has dimensions: 411 rows & 351 columns
Enter resampling option, N)earestneighbor or B)ilinear:  n
Column(x) scale factor:  0.1
```

```
Row(y) scale factor:  0.1
Output has dimensions of 41 rows & 35 columns
New file:  temp2
Description:  *

Command: evaluate
Number of old files:  2
Old file:  eabase
Old file:  temp2
Number of new files:  1
New file 1
 Offsite value:  255
 File name:  ealcov30
Description:  Elm August Water Class 30
Enter evaluate statements.  Use empty line to end.
*  if old2 = 0 then
*    new1 := 0
*  else
*    new1 := old1
*  end;
*  if old1 > 0 and old1 < 255 and old2 > 0 then
*    new1 := old2 * 100 div old1
*  end;
* <cr>
Evaluating.....
```

Step Three: Derive Inventory for Remaining Categories

Repeat the second step as many times as there are categories in the variable. Change the category number assigned to the keyword 'old1' in the **evaluate** command in order to count the occurrence of the next category (e.g. replace "if old1 = 30 then" with "if old1 = 71 then" to count observations of "commercial/industrial" property). Adjust the file name and description in the last **evaluate** command (which calculates the percent importance) to reflect the category just processed (e.g. replace "ealcov30" with "ealcov71" and "Elm August Water Class 30" with "Elm August Com/Ind Class 71").

Step Four: Quit EPPL and Erase Temporary Files

If all categories have been counted for the inventory, leave EPPL7 with the **quit** command. A prompt will appear on the display signaling a succesful return to the operating system. You should at this time erase the three temporary files (TEMP0.EPP, TEMP1.EPP, and TEMP2.EPP) used by EPPL to create the inventory.

```
Command: quit <cr>
exiting ....

C:\SCRATCH> erase temp?.epp
```

REFERENCES

Agricultural Experiment Station, University of Minnesota. 1975. *Soil landscapes & geomorphic regions*, Twin Cities Metropolitan Area sheet, (1:125,000-scale map and summary) Miscellaneous Report 130-1975, St. Paul, MN: University of Minnesota.

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer. 1976. *A land use and land cover classification system for use with remote sensor data*, U.S. Geological Survey Professional Paper 964, Washington, D.C.: U.S. Government Printing Office.

Anderson, Kevin L. 1987. *Bear Creek Surface Water Simulation Modeling Demonstration*, Water Resources Research Center Special Report No. 13, St. Paul, MN: University of Minnesota.

Anderson, Kevin L., J. D. Corbett, C. A. Gersmehl, D. A. Brown, and P. J. Gersmehl. 1986. "Cell size and simulation modeling of water resources at a regional scale," in Bernard J. Nieman, Jr. (editor), *Proceedings*, Urban and Regional Information System Association (URISA), Denver, Volumne I, pp. 203-213.

Brown, Dwight A., and Philip J. Gersmehl (editors). 1987. *File Structure Design and Data Specifications for Water Resources Geographic Information Systems*, Water Resources Research Center Special Report No. 10, St. Paul, MN: University of Minnesota.

Finney, H. R. 1987. Agricultural Extension Service, University of Minnesota, St. Paul, Personal communications.

Gersmehl, Carol A. 1987. "Land Cover Data for a Water Resources GIS," Chapter 7 in D. A. Brown and P. J. Gersmehl (editors), *File Structure Design and Data Specifications for Water Resources Geographic Information Systems*, Water Resources Research Center Special Report No. 10, St. Paul, MN: University of Minnesota.

Gersmehl, Philip J., K. L. Anderson, R. P. Greene, N. P. Dunning, C. A. Gersmehl, and D. A. Brown. 1987. "Hydrologic Classification of Land Cover," Chapter 6 in D. A. Brown and P. J. Gersmehl (editors), *File Structure Design and Data Specifications for Water Resources Geographic Information Systems*, Water Resources Research Center Special Report No. 10, St. Paul, MN: University of Minnesota.

Gersmehl, Philip J., D. A. Brown, and K. L. Anderson. 1987. "File Structure and Cell Size Considerations for a Water Resources GIS," Chapter 2 in D. A. Brown and P. J. Gersmehl (editors), *File Structure Design and Data Specifications for Water Resources Geographic Information Systems*, Water Resources Research Center Special Report No. 10, St. Paul, MN: University of Minnesota.

Gersmehl, Philip J., J. D. Corbett, and R. P. Greene. 1987. "Soil Data for a Water Resources GIS," Chapter 10 in D. A. Brown and P. J. Gersmehl (editors), *File Structure Design and Data Specifications for Water Resources Geographic Information Systems*, Water Resources Research Center Special Report No. 10, St. Paul, MN: University of Minnesota.

Harms, G. F., Soil Conservation Service. 1959. *Soil Survey of Scott County, Minnesota*, St. Paul, MN: USDA.

Hundley, Steven J., Soil Conservation Service. 1983. *Soil Survey of Dakota County, Minnesota*, St. Paul, MN: USDA.

Kent, K. M. 1973. *A method for estimating volume and rate of runoff in small watersheds*, Soil Conservation Service Technical Paper No. 149, Washington, D.C.: USDA.

Lueth, Robert A., Soil Conservation Service. 1974. *Soil Survey of Hennepin County, Minnesota*, St. Paul, MN: USDA.

Metropolitan Council of the Twin Cities. 1977. *Metropolitan Development Guide*, Table 2, Chapter DF, "Development Framework", St. Paul, MN: Metropolitan Council.

Metropolitan Council of the Twin Cities. 1978, 1980, and 1984. *Generalized Land Use, Twin Cities Metropolitan Area*, map, St. Paul, MN: Metropolitan Council.

Minnesota State Planning Agency, Planning Information Center. 1971. *Land use data, 1969*, 40 acre cell resolution, in Minnesota Land Management Information System (MLMIS). Based on map (1:500,000) produced under contract by the University of Minnesota, Minneapolis, MN.

Rust, Richard H. 1986. University of Minnesota, Department of Soil Science, Personal communications.

Rust, Richard H., H. R. Finney, L. D. Hanson, and H. E. Wright, Jr. 1976. "High-altitude Photography in the Development of a Generalized Soil Map," *Soil Science Society of America Journal* 40:405-409.

Soil Conservation Service, USDA. 1979. *Engineering Field Manual for Conservation Practices*, Washington, D.C.: USDA.

Soil Conservation Service, USDA. 1972. *National Engineering Handbook*, Section 4, Hydrology, Washington, D.C.: USDA.

Soil Conservation Service, USDA. 1986. *Urban Hydrology for Small Watersheds*, Technical Release No. 55, Washington, D.C.: USDA.

Soil Survey Staff, Soil Conservation Service. 1975. *Soil Taxonomy*, Agriculture Handbook No. 436, Washington, D.C.: USDA SCS.

U.S. Bureau of the Census. 1983. *County and City Data Book, 1983*, Table B, Washington, D.C.: U.S. Government Printing Office.

Upper Mississippi River Basin Commission. 1977. *Minneapolis-St. Paul Water and Land: Future Prospectives and Plans*, Level B Study Report, Table V-1, Minneapolis, MN: UMRBC.

Young, Robert A., Charles A. Onstad, David D. Bosch, and Wayne P. Anderson. 1985. *AGNPS I, Agricultural Nonpoint Source Pollution Model: A Large Watershed Analysis Tool; A guide for model users*, Draft Copy, Morris, MN: Agricultural Research Service.